Coalition Formation in Political Games

Daron Acemoglu
MIT

Georgy Egorov
Harvard

Konstantin Sonin
New Economic School

September 2006.

Very Very Preliminary and Incomplete.
Please Do Not Circulate without Permission.

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 subcoalitions should be able to secede and become the new ruling coalition. We first present an axiomatic approach that captures these notions and determines a unique self-enforcing ruling coalition. We then construct a simple dynamic game that encompasses these ideas and propose the notion of *sequentially weakly dominant equilibrium* as an equilibrium concept. We prove that this dynamic game generically has a unique sequentially weakly dominant equilibrium, and this equilibrium coincides with a particular type of trembling hand perfect equilibrium. We then show the equivalence of these equilibria to the self-enforcing ruling coalition emerging from the axiomatic approach and also to the core of a related non-transferable utility cooperative game.

The substantive conclusions of our analysis relate to the structure of ruling coalitions. 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 subcoalition of the ruling coalition that commands a majority is self-enforcing. The major insight that emerges from this characterization is that the coalition is made self-enforcing precisely by the failure of its winning subcoalitions being self-enforcing. This is most simply illustrated by the following simple finding: with 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 much 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.

1 Introduction

The central question of political economy is the determination of the collective choices of groups (e.g., Austen-Smith and Banks, 1999). The celebrated Arrow (im)possibility theorem, however, implies that there is relatively little that can be said about collective choices in general environments (Arrow, 1951). This has motivated much of the current political economy literature, which focuses on collective choices under specific institutions (such as legislative bargaining) or under restrictive assumptions on the preferences of individuals making up the group (such as single-peaked preferences).

An alternative approach to collective decisions over the distribution of scarce resources directly starts from the conflict between individuals (or groups) and their unequal power in the process of making collective choices. For example, we may expect individuals with access to guns and resources to be more influential ("politically more powerful"). In this paper, we investigate collective choices in societies with distributional conflict and different distributions of political power among individuals. More specifically, we ask: how does the society consisting of individuals with different degrees of political power decide the allocation of resources? Do more politically-powerful individuals necessarily receive greater weight in collective choices? What general lessons can we draw about the structure of ruling coalitions?

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, for example a pie of size 1, to be distributed among all individuals. We assume that a *ruling coalition* consisting of a subset of the society's members distributes this resource among its members according to their political power. A ruling coalition is defined as a group of individuals that has total power more than $\alpha \in [1/2, 1)$ times the power of all the individuals in society and has no subcoalition that would like to secede and become the new ruling coalition. Loosely speaking, this implies that the ruling coalition is subject to two constraints; a *power constraint* and an *enforcement constraint*, the first requiring the ruling coalition to be powerful enough, and the second imposing that the ruling coalition should not contain any self-enforcing subcoalition. A subcoalition will be self-enforcing, in turn, if its own winning subcoalitions are not self-enforcing. Intuitively, any subcoalition that is 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

[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.

next round of secession (elimination).

One of the simple but interesting implications of these interactions is that generically (in a sense to be made precise below), two-person coalitions, *duumvirates*, cannot be ruling coalitions, but three-person coalitions, *triumvirates*, can be.[2] This result contains many of the key ideas of the paper, and therefore, we start with a simple example that illustrates it.

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, the two-person coalition is not self-enforcing. We therefore say that a two-person 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 $\gamma_B + \gamma_C > \gamma_A > \gamma_B > \gamma_C$. Clearly no two-person coalition is self-enforcing. The lack of self-enforcing subcoalitions of (A, B, C), however, implies that (A, B, C) is itself self-enforcing. To see this, suppose, for example, that a subcoalition of (A, B, C), (B, C) considers seceding from the original coalition. They can do so since $\gamma_B + \gamma_C > \gamma_A$. However, we know from the previous paragraph that the subcoalition (B, C) is itself not self-enforcing, since after this coalition is established, agent B would secede or "eliminate" C. Anticipating this, agent C would not support the subcoalition (B, C). A similar argument applies for all subcoalitions. Moreover, since agent A 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.

Next, consider a society consisting of four individuals and 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 (A, B, C, D). Nevertheless, D is stronger than any two of A, B, C. This implies that any three-person coalition including D would not be self-enforcing. Anticipating this any two of (A, B, C) would resist D's offer to secede and eliminate C. However, (A, B, C) is self-enforcing, thus the

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

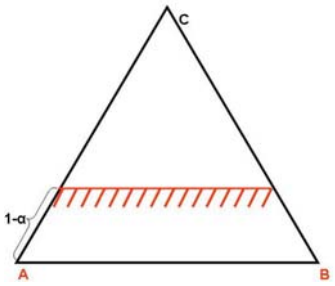

Figure 1: The Power Constraint

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.

Naturally, it is not always the case that the most powerful individual will be eliminated. This can be seen by considering an alternative 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. These two constraints can also be illustrated diagrammatically. Figure 1 depicts the power constraint 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 any generality). The shaded area is the set of all coalitions where the subcoalition (A, B) is winning. The power constraint is parallel to the AB facet of the simplex, which is a general feature. Power constraints are always hyperplanes parallel to a certain facet of the corresponding simplex.

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. Figure 2, for example, 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 always defines a cone (when $N > 3$, it is a quasi-cone, a cone with the vertex being replaced by a facet of the simplex). In the case where $\alpha = 1/2$, this cone becomes a straight line perpendicular to the AB facet.

Using this figure, we can see for which distribution of powers 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 Figure 1. Second, it needs to be self-enforcing, i.e., lie in the cone of enforcement in Figure 2. 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 a two-person coalition

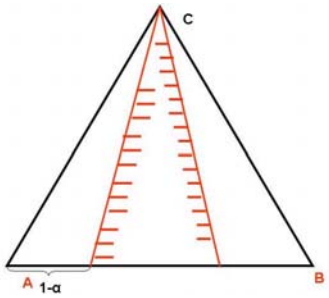

Figure 2: The Enforcement Constraint

cannot become a ruling coalition under majority rule. More generally, given an allocation of powers $\{\gamma_i\}$, a coalition X can only threaten the stability of the allocation if $\{\gamma_i\}$ if it satisfies the power constraint for coalition X (that is, X must be winning) and it lies within the enforcement cone (that is, X must be self-enforcing).

Our first result is that an axiomatic approach to the determination of the ruling coalition using these two notions is sufficient to determine a unique ruling coalition in generic games. We achieve this by defining a mapping from the set of coalitions of the society into itself and imposing some minimal conditions capturing the power and the enforcement constraints (as well as an individual rationality type condition). The coalition determined by this map applied to the entire set of players gives the *self-enforcing ruling coalition*.

That this axiomatic approach and the notion of self-enforcing 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 want to engage in further elimination, emerges. This ultimate ruling coalition divides the scarce resource among its members according to their power. The important assumption here is that 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.[3] We then establish the generic existence and uniqueness of a sequentially weakly

[3]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, which was the first to predict such a relationship)".

dominant equilibrium and a Markov trembling hand perfect equilibrium of this dynamic game,[4] and we show that the equilibrium outcomes coincide with the self-enforcing ruling coalition derived from the axiomatic approach. Finally, we also 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 publications that give high payoff to their members.

All of these approaches give the same solution because they capture the same salient features of the process of collective decision-making. First, the distribution of power matters for the resolution of conflict among the members of the society. Second, coalitions between different individuals emerge in equilibrium. Third, a more powerful individual need not obtain a greater share of resources in society, since the distribution of resources will be determined in equilibrium, depending on what types of coalitions form.

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

1. There always exists a self-enforcing ruling coalition and can be computed by induction.

2. 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.

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

4. Nevertheless, self-enforcing ruling coalitions are continuous in the distribution of power across individuals in the sense that itself-enforcing ruling coalition remains so when the powers of the players are perturbed by a small amount.

5. 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, again under majority rule, coalitions where members have roughly the same power exist only when the coalition's size is $2^k - 1$ where k is an integer.

[4] 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.

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

7. 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.

Our paper is related to a number of different literatures. The first is the social choice literature (e.g., Austen-Smith and Banks, 1999). The difficulty of determining the social welfare function of a society highlighted by Arrow's theorem is related to the fact that the core of the game defined over the allocation of resources is empty. As we establish below, our approach is equivalent to looking at a weaker notion than the core, whereby only "self-enforcing" coalitions are allowed 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.

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 bargaining outcomes 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 Shunik (1954) on power struggles in committees and the paper by Aumann and Kurz (1977), which looks at the Shapley value of a bargaining game in order to determine the distribution of resources in the society. 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., Hart and Kurz, 1983, Aumann and Myerson, 1988, Greenberg and Weber, 1993, Chwe, 1994, Bloch, 1996, Ray and Vohra, 1999, Konishi and Ray, 2001, Maskin, 2003).[5] The most important difference between our approach and the previous literature on coalition formation is that, motivated by political settings, we assume that the majority (or supermajority) of the members of the society

[5] Like some of these papers, our approach can be situated within the "Nash program" since our axiomatic approach is supported by an explicit extensive form game (Nash, 1953). See Serrano (2004) for a recent survey of work on the Nash program.

can impose their will on those players who are not a part of the majority.[6] This feature both changes the nature of the game and also introduces "negative externalities" as opposed to the positive externalities and free-rider problems on which the previous literature focuses (see, for example, Ray and Vohra, 1999, Maskin, 2003). A second important difference is that most of these works assume the possibility of binding commitments (see again Ray and Vohra, 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. Osborne and Rubinstein (1994, chapter 14) gives a comprehensive discussion 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 the salient features of political decision-making. Section 3 provides our axiomatic treatment of this game. It introduces the concept of self-enforcing ruling coalition and proves its 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 self-enforcing 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 unique core allocation of this game and the self-enforcing ruling coalition. 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 Appendix contains all the proofs not provided in the text as well as a number of examples to further motivate some of our equilibrium concepts.

2 The Political Game

We now describe the environment for collective decision-making. Consider a society consisting of a finite set of individuals $N = \{1, 2, ..., |N|\}$. The society has a resource of size 1 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

[6] 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.

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 how differences in the powers of individuals (or groups) map into political decisions. For this reason, we assume that each individual $i \in N$ is endowed with political *power* $\gamma_i \in \mathbb{R}_{++}$ (where $\mathbb{R}_{++} = \mathbb{R}_+ \setminus \{0\}$). For every set X denote the set of its subsets by $P(X)$. Any element $X \in P(N)$ is called a *coalition*. The value

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

is called the *power* of coalition X

We assume that collective decisions require a (super)majority. 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 $X \in P(N)$ and $Y \in P(N)$. The coalition Y is* winning within X *if*

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

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

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 winning coalition Y 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, we define an abstract political game as $\Gamma = \left(N, \{\gamma_i\}_{i \in N}, \alpha\right)$. 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, 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 assume that if X is a ruling coalition, then it distributes the scarce resource among its members according to their power. In particular, if $X \subset N$ is a ruling coalition, then the share of the resource received by any player $i \in N$ is given by

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

Evidently, for any $X \subset N$,

$$\sum_{i \in N} w_i(X) = 1.$$

The assumption that a ruling coalition decides the distribution of resources is without any loss of generality. The assumption that resources are distributed according to the power of the members of the ruling coalition is also not very restrictive; we will focus on pure strategies and all players have strictly increasing preferences over their share, thus any sharing rule within the ruling coalition that gives weakly greater shares to more powerful members will lead to similar results.

The more important assumption introduced so far is that a coalition cannot commit to a distribution of resources among its members. For example, a coalition consisting of two individuals with powers 1 and 10 cannot commit to giving the entire resource to the first individual 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 processes; political decisions are made 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 3).

Since equation (2) uniquely defines the division of the resource given the ruling coalition, the outcome of any game can be represented by its ruling coalition alone. More formally, let $\mathcal{G}$ be the set of all possible games of the form $\Gamma = \left(N, \{\gamma_i\}_{i\in N}, \alpha\right)$. The outcome function determines a subset of N as the ruling coalition for any game $\left(N, \{\gamma_i\}_{i\in N}, \alpha\right)$, i.e.,

$$\Phi_a : \mathcal{G} \to P(N).$$

In this paper, we are interested with the properties of this outcome function. In particular, 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., inclusive).

In the text, we focus on games that satisfy the following genericity assumption (see Appendix B for generalizations).

Assumption 1 *Numbers* $\{\gamma_i\}_{i\in N}$ *are* generic *in the sense that there does not exist coalitions* X *and* Y *of* N *such that* $X \neq Y$, $\gamma_X = \gamma_Y$ *or* $\alpha\gamma_X = (1-\alpha)\gamma_Y$.

Intuitively, this assumption rules out distributions of powers among individuals such that two different coalitions will have exactly the same total power or a ratio $\alpha/(1-\alpha)$ of each other's power (these two conditions are clearly the same when $\alpha = 1/2$). The reason for ruling out such coalitions is that, when they exist, they will lead to a type of non-uniqueness, which leads to uninteresting technical problems. Notice that this assumption is without much loss of generality since for any society N the set of vectors of $(\gamma_1, ..., \gamma_N) \in \mathbb{R}^{|N|}_{++}$ that fail to satisfy Assumption 1 are of Lebesgue

measure 0 in $\mathbb{R}_{++}^{|N|}$ (in fact, it has Lebesgue measure 0 in any subset of $\mathbb{R}_{++}^{|N|}$ with nonempty interior). 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 focus is to determine some general features of the outcome function Φ_a defined above when we impose certain natural axioms. Our analysis and axioms are motivated by the discussion in the Introduction, which suggested two constraints; the power constraint and the enforcement constraint. In particular, we would like the outcome function to pick a winning coalition (according to Definition 1) and to be able to withstand challenges from coalitions that satisfy the enforcement constraint (i.e., from coalitions that are self-enforcing). We will call such a coalition a *self-enforcing ruling coalition*.

The basis of our axiomatic treatment is to fix a game $\Gamma = \left(N, \{\gamma_i\}_{i \in N}, \alpha\right)$ with $\alpha \in [1/2, 1)$ and define a selection mapping ϕ_Γ, which selects a subcoalition Y of any coalition X of N as a "self-enforcing" winning coalition within X. To simplify notation, we drop the dependence on Γ and refer to this mapping as ϕ whenever this will cause no confusion. Formally, for a given $\Gamma = \left(N, \{\gamma_i\}_{i \in N}, \alpha\right)$, we have

$$\phi : P(N) \to P(N).$$

In the spirit of the power and the enforcement constraints, we adopt the following axioms on ϕ. Fix a game $\Gamma = \left(N, \{\gamma_i\}_{i \in N}, \alpha\right)$. Then we impose:

Axiom 1 ***(Power)*** *For any $X \in P(N)$ and $Y \in P(X)$, $\phi(X) = Y$ implies that $\gamma_Y \geq \alpha \gamma_X$.*

Axiom 2 ***(Enforcement)*** *For any $X \in P(N)$ and $Y \in P(X)$, $\phi(X) = Y$ implies that $\phi(Y) = Y$.*

Axiom 3 ***(Individual Rationality)*** *For any $X \in P(N)$ and $Y \in P(X)$, $\phi(X) = Y$ implies that $\gamma_Y \leq \gamma_Z$ for all $Z \in P(X)$ such that $\gamma_Z \geq \alpha \gamma_X$ and $\phi(Z) = Z$.*

We say that a coalition $X \in P(N)$ is a *self-enforcing ruling coalition* if $\phi(N) = X$.

All three axioms are natural. The first one, the power axiom, requires that the winning coalition has sufficient power according to Definition 1. The second axiom, the enforcement axiom, simply states that a self-enforcing winning coalition should be self-enforcing, i.e., it should select itself. The final axiom requires that if there are multiple self-enforcing winning coalitions, the one with the minimal power should be selected. This is an individual rationality type axiom. To see this,

note that since $\alpha \geq 1/2$, if there exist two self-enforcing winning coalitions Y and Z, they cannot be disjoint, i.e., $Y \cap Z \neq \varnothing$. Given the division rule in (2), the common members of these two coalition would be better off with whichever coalition has less total power, and thus individual rationality dictates that they should join the least powerful self-enforcing winning) coalition. Axiom 3 imposes this requirement.

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

Theorem 1 *Consider a game* $\Gamma = \left(N, \{\gamma_i\}_{i \in N}, \alpha\right)$ *with* $\alpha \in [1/2, 1)$ *and suppose that Assumption 1 holds. Then there exists a unique mapping* ϕ *that satisfies Axioms 1-3. Moreover,* ϕ *is single-valued.*

Proof. The proof is by induction. Start with a singleton $\{i\}$. Clearly $\phi(\{i\}) = \{i\}$ satisfies Axioms 1-3, and $\phi(\{i\}) = \varnothing$ fails to do so. Thus the mapping ϕ is uniquely defined and single-valued for X such that $|X| = 1$. Next suppose that ϕ is uniquely determined for all Z such that $|Z| \leq n$ and consider X, where $|X| = n+1$. Consider all $Y \subset X$ such that $\gamma_Y \geq \alpha\gamma_X$, and denote the set of all such Y's by X_Y. By definition, $\phi(Y)$ is determined for all $Y \in X_Y$. If there exists $Y_1, ..., Y_k \in X_Y$ such that $\phi(Y_j) = Y_j$ for $j = 1, ..., k$, then by Axioms 1-3 and Assumption 1 $\phi(X)$ is uniquely determined as $\phi(X) = Y_{k'}$ such that $\gamma_{Y_{k'}} < \gamma_{Y_j}$ for all $j = 1, ..., k$, $j \neq k'$ (since by Assumption 1, that cannot exist two subsets of X_Y with the same power). Suppose that X_Y is empty. Then from Axioms 1-2, $\phi(X) = X$. This implies that ϕ is uniquely defined and single-value for all sets X with $|X| = n+1$. This completes the proof of the induction step and thus the proof of the theorem. ■

At first, Axioms 1-3 may appear relatively mild. Nevertheless, they are strong enough to pin down a unique mapping ϕ, which is also single-valued. The more substantive part of Theorem 1 is the existence and uniqueness of the mapping ϕ. That it is single-valued follows in view of Axiom 3 and Assumption 1.

Motivated by Axioms 1-3 and Theorem 1, throughout we refer to coalitions X such that $\phi(X) = X$ as *self-enforcing* coalitions.

The fact that ϕ is single-valued implies the following corollary.

Corollary 1 *Under the assumptions of Theorem 1, there exists a unique self-enforcing ruling coalition.*

Proof. This immediately follows since $\phi(N)$ exists and is unique. ■

Theorem 1 and Corollary 1 are stated under Assumption 1. The mapping ϕ is still well defined when this assumption is relaxed, but it is no longer single-valued. Theorem 6 in Appendix C deals

with this case. To simplify the exposition, throughout the text we focus on generic games where Assumption 1 holds.

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

Example 1 (continued) For continuity with that example, let the players be denoted by A, B and C (rather than 1, 2 and 3) and suppose that $\alpha = 1/2$. For any $\gamma_A < \gamma_B < \gamma_C$ that satisfy Assumption 1 and $\gamma_A + \gamma_B > \gamma_C$, $\phi(\{A, B, C\}) = \{A, B, C\}$. To see this, it suffices that 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 2. 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 1. Since $\phi(\{A, B, C\})$ exists by Theorem 1, it must be that $\phi(\{A, B, C\}) = \{A, B, C\}$. We can also see where this line of argument would go wrong if Assumption 1 were not satisfied. In that case, we could have $\gamma_A = \gamma_B$, and $\phi(\{A, B\}) = \{A, B\}$. As long as $\gamma_C > \gamma_A = \gamma_B$, $\phi(\{A, B, C\})$ would still be well-defined and single-valued. However, if we also had $\gamma_A = \gamma_B = \gamma_C$, $\phi(\{A, B, C\})$ could be any two-person coalition, and thus the mapping ϕ would no longer be single valued.

We next characterize the ϕ mapping and determine the structure and properties of self-enforcing ruling coalitions. Before doing this, however, we will present a dynamic game and then a cooperative game, which will further justify our axiomatic approach. Our analysis of these games will also provide us with an inductive way of determining $\phi(X)$ for any set X.

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, the unique equilibrium (under Assumption 1) will coincide with the self-enforcing ruling coalition defined in the previous section.

4.1 The Basic Game Form

Consider a society N consisting of a finite number of individuals, with a distribution of power $\{\gamma_i\}_{i \in N}$, and an institutional rule $\alpha \in [1/2, 1)$. We will denote the corresponding extensive-form game by $\hat{\Gamma} = \left(N, \{\gamma_i\}_{i \in N}, \alpha\right)$. Note that $\hat{\Gamma}$ is different from Γ defined in the previous section, since it refers to the extensive form game described next.

Let $\varepsilon > 0$ be a small number. Then the extensive form of the game $\hat{\Gamma}$ is as follows.

1. At each stage, $j = 0, 1, ...$, 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 $\text{Yes}\{X_{j_q}\}$ be the subset of X_{j_q} voting in favor of this proposal. 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 is 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 as given by (2).

In other words, 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 (except 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

$$U_i = w_i(N_{\bar{j}}) - \varepsilon \sum_{1 \leq j \leq \bar{j}} I_{N_{\bar{j}}}(i), \tag{3}$$

where $I_X(\cdot)$ is the indicator (characteristic) function of set X. This payoff function captures the fact 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 (3)).

With a slight abuse of terminology, we refer to j above as "the stage of voting," so that if the ultimate ruling coalition is reached when $j = 0$, we say that the game ended in the first stage of voting.

Without loss of any 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.

Throughout, we focus on the case where ε is arbitrarily small. The cost ε can be interpreted as a cost of eliminating some of the players from the coalition or as an organizational cost that individuals have to pay each time a new coalition is formed. Its role for us is to rule out some unintuitive equilibria that arise in dynamic voting games. Example 3 below illustrates the types of equilibria that arise when $\varepsilon = 0$.

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 non-increasing sequence over a compact set and necessarily converges, determining an ultimate ruling coalition. Consequently, the extensive-form game $\hat{\Gamma}$ necessarily has a subgame perfect Nash equilibrium (SPNE). However, as the next example shows, there may be many SPNEs, some of them unintuitive.

Example 2 Consider $N = \{1,2,3,4\}$, with $\gamma_1 = 2$,$\gamma_2 = 4$, $\gamma_3 = 7$ and $\gamma_4 = 10$, and suppose that $\alpha = 1/2$. From Theorem 1, it can be seen that $\phi(\{1,2,3,4\}) = \{1,2,3\}$. Now suppose that nature picks player 1 as the initial proposer, and this player proposes $X_1 = \{1,2,3\}$. It may appear natural to imagine that this coalition will receive the majority of the votes. Nevertheless, for reasons that are familiar from voting games more generally, this may not be the case. For example, all four players voting against this proposal constitutes a best response for each player, since no single player can change its vote and affect the voting outcome. Consequently, both $\{1,2,3,4\}$ and $\{1,2,3\}$ can emerge as subgame perfect equilibrium URCs, even though the former is not a reasonable outcome, since $1,2$ and 3 have enough votes to eliminate 4. In voting games, equilibria like the one involving $\{1,2,3,4\}$ as the URC are eliminated by focusing on weakly dominant strategies. In particular, voting in favor of X_1 is a weakly dominant strategy for players 1, 2 and 3. However, it can be verified that in a multi-stage voting game voting against coalitions like X_1 need not be a weakly dominated strategy (see Example 4 in Appendix A). For this reason, in the next section, we introduce the concept of sequentially weakly dominant equilibrium.

4.2 Sequentially Weakly Dominant Equilibria

In this subsection, we introduce the notion of *Sequential Weakly Dominant Equilibrium* inductively for finite games. To define this solution concept, we first 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

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

with $a_t^i \in A_t^i$ and

$$a^i \in A^i \equiv \prod_{t=1}^{T} A_t^i.$$

Let $h^t = \left(h_1, ..., h^t\right)$ be the history of play up to stage t, where $h_t = \left(a_t^1, ..., a_t^n\right)$,with $h_t \in H_t$. We denote the set of all potential histories up to date t by

$$H^t \equiv \prod_{s=1}^{t} H_s.$$

Let t-continuation action profiles be

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

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

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

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. The payoff functions for the players depend only on actions, i.e.,

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

We also define the restriction of the payoff function u^i to a continuation play $\left(a^{1,t}, ...a^{n,t}\right)$ as

$$u^i\left(a^{1,-t}, ...a^{n,-t} \ \vdots \ a_t^1, ...a_t^n \ \vdots \ a^{1,t+1}, ...a^{n,t+1}\right).$$

In words, this specifies the utility to player i from having played action profile $\left(a^{1,-t}, ...a^{n,-t}\right)$ up to and including time $t-1$, playing the action profile $\left(a_t^1, ...a^n\right)$ at time t and being restricted to the action profile $a^{1,t}, ...a^{n,t}$ from t onwards. Symmetrically, this payoff function can also be read as the utility from continuation action profile $\left(a^{1,t+1}, ...a^{n,t+1}\right)$ given that up to time t, the play has consisted of the action profile $\left(a^{1,-t}, ...a^{n,-t} \ \vdots \ a_t^1, ...a_t^n\right)$.

A (possibly mixed) strategy for player i is

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

where $\Delta(X)$ denotes the set of probability distributions defined over the set X.

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, i.e.,

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

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, i.e.,

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

where $H^T \backslash H^{t-1}$ denotes all histories starting from time t onwards.

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 it uses the continuation strategy $\sigma^{i,t}$ and other players use $\sigma^{-i,t}$. We also use the notation $u^i\left(\sigma^{1,t}, ...\sigma^{n,t} \ \vdots \ \sigma^{1,t+1}, ...\sigma^{n,t+1} \mid h^{t-1}\right)$ as the payoff from strategy profile $\left(\sigma^{1,t}, ...\sigma^{n,t}\right)$ at time t restricted to the continuation strategy profile $\left(\sigma^{1,t+1}, ...\sigma^{n,t+1}\right)$ 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 it 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 2 *A strategy profile* $\left(\hat{\sigma}^1, ..., \hat{\sigma}^n\right)$ *is a* Nash Equilibrium *if and only if*

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

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

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

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 4 *We say that* $\left(\hat{a}^1, ...\hat{a}^n\right)$ *is a weakly dominant equilibrium if*

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

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. At weakly dominant strategy equilibrium in this last stage of the game is defined similar to Definition 4.

Definition 5 *There exists a h^{T-1}-weakly dominant equilibrium if there exists $\left(\hat{\sigma}^{1,T},...,\hat{\sigma}^{n,T}\right)$ such that*

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

Now inductively, we can define sequentially weakly dominant strategy equilibria.

Definition 6 *There exists a h^{t-1}-sequentially weakly dominant equilibrium for $t < T$ if there exists a h^t-sequentially weakly dominant equilibrium given by $\left(\hat{\sigma}^{1,t+1},...,\hat{\sigma}^{N,t+1}\right)$ and*

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

In words, we first hypothesize that there exists a h^t-sequentially weakly dominant equilibrium, and impose that this 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 *A finite game has a* Sequentially Weakly Dominant Equilibrium (SWDE) *if it has a h^0-sequentially weakly dominant equilibrium.*

We refer to the strategy profile played along the equilibrium path of this sequentially weakly dominant equilibrium as the sequentially weakly dominant equilibrium strategy profile.

With this terminology, we can also introduce the notion of Markov Trembling Hand Perfect Equilibria.

Definition 8 *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

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

Let M be an index set. We also define:

Definition 9 *We say that a strategy profile* $\left(\hat{\sigma}^1, ..., \hat{\sigma}^n\right)$ *is* Markov Trembling-Hand Perfect Equilibrium (MTHPE) *if there exists a sequence of totally mixed Markovian strategy profiles in the agent-normal form* $\left\{\left(\hat{\sigma}^1(m), ..., \hat{\sigma}^n(m)\right)\right\}_{m\in M}$ *such that* $\left(\hat{\sigma}^1(m), ..., \hat{\sigma}^n(m)\right) \to \left(\hat{\sigma}^1, ..., \hat{\sigma}^n\right)$ *and*

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

Note that MTHPE is defined directly on the agent-normal 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 MTHPE for game $\hat{\Gamma}$ and show their equivalence.

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 $\hat{\Gamma} = \left(N, \{\gamma_i\}_{i\in N}, \alpha\right)$, there is a corresponding abstract game $\Gamma = \left(N, \{\gamma_i\}_{i\in N}, \alpha\right)$. Recall that $\mathcal{G}$ denotes the set of all such abstract games, and can be interchangeably used to denote the set of all extensive form games as described in this section. Our axiomatic approach in Section 3 specified a mapping $\Phi_a : \mathcal{G} \to P(N)$, which determined the self-enforcing ruling coalition for each game $\Gamma \in \mathcal{G}$. In particular, Theorem 1 shows that this can be represented as $\phi_\Gamma(N)$ for a well-defined single-valued mapping ϕ_Γ. Similarly, we can think of an outcome mapping $\hat{\phi}_{\hat{\Gamma}}$, which determines an ultimate ruling coalition $\hat{\phi}_{\hat{\Gamma}}(N) \in P(N)$ for each extensive form game $\hat{\Gamma} \in \mathcal{G}$. In particular, $\hat{\phi}_{\hat{\Gamma}}(N)$ would designate the URC that arises as the SWDE of this extensive-form game. Our main result in this section will be the equivalence result that for any $\hat{\Gamma} = \left(N, \{\gamma_i\}_{i\in N}, \alpha\right)$ and $\Gamma = \left(N, \{\gamma_i\}_{i\in N}, \alpha\right)$,

$$\hat{\phi}_{\hat{\Gamma}}(N) = \phi_\Gamma(N).$$

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

Theorem 2 *Any extensive-form game* $\hat{\Gamma} = \left(N, \{\gamma_i\}_{i\in N}, \alpha\right)$ *has at least one pure strategy SWDE. Moreover, suppose that Assumption 1 holds. Then in any pure strategy SWDE, the ultimate ruling coalition (URC) is reached after one stage of voting, is given by* $\phi(N)$ *as defined in Theorem 1, and the payoff of each* $i \in N$ *is given by* $U_i(N) = w_i(\phi(N)) - \varepsilon I_{\{i\in\phi(N)\}} I_{\{\phi(N)\neq N\}}$.

Proof. See Appendix B. ■

This theorem establishes two important results. First, a pure-strategy SWDE exists for any game $\hat{\Gamma} = \left(N, \{\gamma_i\}_{i\in N}, \alpha\right)$ and the URC is reached in the first stage of voting. Moreover, this pure-strategy SWDE is independent of the moves by nature (i.e., of the exact ordering of proposals chosen

by nature). The existence result 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. Second, the SWDE ultimate ruling coalition coincides with $\phi(N)$, that is, with the self-enforcing ruling coalition of $\Gamma = \left(N, \{\gamma_i\}_{i\in N}, \alpha\right)$, which was derived axiomatically in Section 3. Moreover,Given this equivalence, throughout, we use the terms URC and self-enforcing ruling coalition interchangeably.

Neither Theorem 1 nor Theorem 2 provide a characterization of the self-enforcing ruling coalition, $\phi(N)$. This ruling coalition can be determined inductively. This is done in the next proposition.

Proposition 1 *Consider a game* $\Gamma = \left(N, \{\gamma_i\}_{i\in N}, \alpha\right)$ *that satisfies Assumption 1. Then:*

1. *For any* $X \subset N$*, if* $\phi(X) = X$*, then there does not exist* $Y \subset X$ *such that* $\gamma_Y > \alpha\gamma_X$ *and* $\phi(Y) = Y$*.*

2. *Denote the set of* $X \subset N$ *such that* $|X| = k$ *by* N^k*. The self-enforcing ruling coalition,* $\phi(N)$*, can be computed inductively as follows. Define the order* $\succ$ *over sets, such that* $X' \succ X''$ *if and only if* $\gamma_{X'} > \gamma_{X''}$ *and the* min *operator over sets according to this order. For* $k = 1, ..., |N|$*, let* $\mathbf{X}^k = X^k \cup \left\{Y \in P\left(X^k\right) : \gamma_Y > \alpha\gamma_{X^k} \text{ and } \phi(Y) = Y\right\}$*. Then*

$$\phi\left(X^k\right) = \min\left\{X^s : X^s \in \mathbf{X}^k\right\}.$$

Part 1 of this proposition follows by definition. Part 2 gives an algorithm that can be used easily to compute the self-enforcing ruling correlation. The set $\mathbf{X}^k$ is defined to include the coalition X^k itself, so that if the set $\left\{Y \in P\left(X^k\right) : \gamma_Y > \alpha\gamma_{X^k} \text{ and } \phi(Y) = Y\right\}$ is empty, $\phi\left(X^k\right) = X^k$ in line with Part 1 of the proposition. For example, clearly any set X^1 in N^1 satisfies $\phi\left(X^1\right) = X^1$. Then consider N^2 and find all self-enforcing coalitions of size 2. For example, when $\alpha = 1/2$, given any distribution of power that satisfies Assumption 1, $\phi\left(X^2\right) \neq X^2$ for any $X^2 \subset N^2$. Then consider the subsets of N^3. Since there exist no $X^2 \subset N^2$ with $\phi\left(X^2\right) = X^2$, this implies that $\phi\left(X^3\right) = X^3$ or $\phi\left(X^3\right) = \{i\}$ for some $i \in X^3$ for any $X^3 \subset N^3$. Proceeding inductively in this manner, we can compute $\phi(X)$ for any $X \subset N$ including for $X = N$, which gives us the self-enforcing ruling coalition.

Finally, the next example shows that when $\varepsilon = 0$, there may exist some unintuitive SWDEs, motivating our choice of payoff function with a small positive value of ε.

Example 3 Let $\alpha = 1/2$, $|N| = 4$, and let players' strengths be $2, 4, 7, 10$. Assume that some elimination in which 2 survives has taken place. Then it is straightforward to check that the

strongest player among those who survive will have enough power to eliminate the rest, and will certainly do it in equilibrium. We next show that different ultimate ruling coalitions may emerge in equilibrium when $\varepsilon = 0$.

(i) Suppose 10 is the first to propose and proposes $N_1 = (10, 2)$. Suppose also that 10 and 2 vote for the proposal, and after 4 and 7 are eliminated, 10 eliminates 2. If, however, this proposal is rejected, then 4 and 7 make proposal $(4, 7, 10)$ and these three vote for it; as for 2, it proposes the grand coalition and this offer is rejected by 4, 7, and 10. It is easy to check that this is an equilibrium. Indeed, 2 is eliminated sooner or later, and it is a best response for him to vote for 10's proposal. So, 10 may be the only surviving player.

(ii) Now suppose that 2 and 4 were the first to make proposals, and their proposals (the grand coalition and $(4, 7, 10)$ were rejected. Then, 7's proposal $(4, 7, 10)$ may be accepted in equilibrium: indeed, 10 knows that his proposal of $(10, 2)$ (or any other proposal which will eventually make him the only surviving player) will be rejected so that the grand coalition will form. Given this, it is Advest response for 2 to propose the grand coalition and for 4 to propose $(4, 7, 10)$, and for other players to reject it. Therefore, $(4, 7, 10)$ will emerge in equilibrium.

This example illustrates that $\varepsilon = 0$ may lead to a number of unappealing results. First, the outcome may depend on the order of proposals. Second, players that will be eliminated (like 2 in this example) may 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.

4.4 Markov Trembling Hand Perfect Equilibria and Agenda-Setting Games

In this section, we show that for our extensive form game the SWDE also coincides with the MTHPE as defined above. Moreover, we show the existence of a unique MTHPE, which is also an SWDE, for a more general class of political games, which we refer to as *agenda-setting games*.[7]

We will also see that the MTHPEs of our extensive-form game of coalition formation lead to the same ultimate ruling coalitions as the SWDEs of the same game. This is not a general result, since these two equilibrium concepts do not coincide. First, a MTHPE always exists, while SWDE may not. Second, there may exist SWDEs that are not MTHPE (see Appendix A).

Our main result here is the following.

[7] Another trembling hand refinement used in the literature, *truly perfect equilibrium*, is stronger than our notion of MTHPE. It 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. However, this equilibrium concept fails to exist in many games, including in our extensive-form game of coalition formation (except in some special cases).

Theorem 3 *Any extensive-form game* $\hat{\Gamma} = \left(N, \{\gamma_i\}_{i \in N}, \alpha\right)$ *has at least one pure strategy MTHPE. Moreover, suppose that Assumption 1 holds. Then any pure strategy MTHPE coincides with the SWDE, and has an ultimate ruling coalition (URC) given by* $\phi(N)$ *as defined in Theorem 1, the URC is reached after one stage of voting, and the payoff of each* $i \in N$ *is given by* $U_i(N) = w_i(\phi(N)) - \varepsilon I_{\{i \in \phi(N)\}} I_{\{\phi(N) \neq N\}}$.

Proof. Consider a perturbed game where each player $i \in N$ plays a mixed Markovian strategy assigning probability $\eta_i^k > 0$ to each of its finite number of actions. By the standard fixed theorem argument, this perturbed game has a Nash equilibrium, and the correspondence determining the Nash equilibrium has closed graph. Taking the limit as $\eta_i^k \to 0$ for all i and k gives an equilibrium by the closed graph property and is, by definition, a MTHPE. Next take any strategy profile σ that forms a MTHPE. By Theorem 4 below σ is also a SWDE, and the conclusion follows from Theorem 2. ■

We now define general agenda-setting games, which include most voting games as a special case, and establish the existence of pure strategy MTHPE and SWDE for these games.

Definition 10 *A finite perfect-information game* Γ *in extensive form with a set of players* $N \cup \{Nature\}$ *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* $N \backslash X$ *have no action at this stage;*
   (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* $i \in X$*, holding other players' actions 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 is one of the agents makes a proposal and others vote in favor or against this proposal is an agenda-setting game. Clearly, our a dynamic game here is an agenda-setting game.

The following theorem shows that, while MTHPE and SWDE are not subsets of each other in general, but for agenda-setting games an MTHPE is always a SWDE.

Theorem 4 *1. If Γ is an agenda-setting game, any MTHPE is a SWDE.*

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

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

Proof. See Appendix B. ■

An immediate corollary of Part 1 of this theorem is that a SWDE always exists in agenda-setting games (since a MTHPE always exists and is a SWDE). Another direct corollary of Part 1 is Theorem 3, which establishes the equivalence of these two equilibrium concepts in our game, though Part 3 shows that there existed in the-sitting games where some SWDE are not MTHPE (thus showing that the latter concept is stronger).

5 A Cooperative Game

In this section, we present a non-transferable utility cooperative game and establish that the (generically) unique allocation in the strong core coincides with the results in the previous two sections.[8] This exercise is useful since it links our equilibrium concept to those in the cooperative game theory literature, and also provides another justification for our axiomatic approach.

An non-transferable utility cooperative game is represented by $\Gamma_N = \left(N, \{\gamma_i\}_{i\in N}, \alpha, v_N(\cdot)\right)$, where $v_N : P(N) \to P\left(\mathbb{R}_+^{|N|}\right)$ 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 first define the mapping v_N inductively. First, define the set of feasible allocations (for any vector $x \in \mathbb{R}_+^{|N|}$, denote its ith component by x^i and its projection on components from X by x^X).

Definition 11 *A vector $x \in \mathbb{R}_+^{|N|}$ is a* feasible allocation *if either*

1. *$x^i = 0$ for all $i \in N$, or*

2. *$x^i = \gamma_i/\gamma_N$ for all $i \in N$, or*

3. *there exists a subcoalition $Y \subset N$, $Y \neq N$, such that x^Y is the core allocation for $\Gamma_Y = \left(Y, \{\gamma_i\}_{i\in Y}, \alpha, v_Y(\cdot)\right)$ (determined by induction), while $x^i = 0$ for all $i \notin Y$.*

[8] By "strong core," we referr to an allocation that cannot be strictly improved upon for all members of a blocking coalition; see Definition 12. To reduce terminology, refer to this as the "core".

This definition states that feasible allocations include those where all individuals receive zero payoff, or those in which all individuals in the society share the resource according to their powers (which could be referred to as the "status quo" publication), or a coalition Y distributes the resource among its members. This definition makes reference to core allocations, which will be defined below (thus the set of feasible allocations and core allocations are determined inductively).

Now define the mapping v_N as follows:

$$v_N(X) = \begin{cases} \left\{x \in \mathbb{R}_+^{|N|} \mid x \text{ is feasible}\right\} & \text{if } \gamma_X > \alpha\gamma_N \\ \left\{x \in \mathbb{R}_+^{|N|} \mid x^i = 0 \ \forall\ i \in N\right\} & \\ \text{or } \left\{x \in \mathbb{R}_+^{|N|} \mid x^i = \gamma_i/\gamma_N \ \forall\ i \in N\right\} & \text{if } (1-\alpha)\gamma_N \leq \gamma_X \leq \alpha\gamma_N \\ \left\{x \in \mathbb{R}_+^{|N|} \mid x^i = 0 \ \forall\ i \in N\right\} & \text{if } \gamma_X < (1-\alpha)\gamma_N \end{cases} \tag{4}$$

where notice that $v_N(X)$ is a a subset of $\mathbb{R}_+^{|N|}$, meaning that it consists of a set of vectors in $\mathbb{R}_+^{|N|}$.

In words, the set of payoff allocations that can be enforced by a coalition X, $v_N(X)$, include any feasible allocation if the coalition is winning within N according to Definition 1 (the first term in (4)). If, on the other hand, the complement of X, $N\backslash X$, is winning, then coalition X can only enforce zero payoff to all (the third term in (4)). When $\alpha = 1/2$, these two are the only possibilities. However, for $\alpha > 1/2$, neither a coalition nor its complement may be winning. The second term in (4) then states that in this case coalition X can enforce either zero payoff to all or the division of the resource among all players according to their powers.

This payoff mapping captures the idea that a coalition which is not the majority can at best block what other coalitions can do and thus implements the division of the scarce resource according to the power of all individuals in the society. A coalition that forms a majority can implement any feasible allocation.

We denote the core allocations for a non-transferable utility game $\Gamma_N = \left(N, \{\gamma_i\}_{i\in N}, \alpha, v_N(\cdot)\right)$ by $C(\Gamma_N) \subset \mathbb{R}_+^{|N|}$ and define this in the standard way.

Definition 12 *A vector $x \in \mathbb{R}_+^{|N|}$ is in the (strong) core for the game $\Gamma_N = \left(N, \{\gamma_i\}_{i\in N}, \alpha, v_N(\cdot)\right)$, i.e., $x \in C(\Gamma_N)$, if and only if it is a feasible allocation and there exists no $Z \in P(N)$ and no $z \in v_N(Z)$ with $z^i > x^i$ for all $i \in Z$.*

Notice that Definition 11 rules out arbitrary distributions of the scarce resource as feasible. For example, in a society consisting of three individuals with powers $(5, 9, 11)$, the resources being divided between the three individuals which shares $(5/25, 9/25, 11/25)$ is a feasible allocation. To

check whether this allocation is in the core, we need to see whether there exists another coalition that implements a feasible allocation improving the payoff to each of its members. Consider, for example, the coalition $(5, 9)$. Since this coalition has a winning majority within $(5, 9, 11)$, (4) states that it can implement $(5/25, 9/25, 11/25)$, but this is exactly the same as the candidate core allocation. In addition, it may be able to implement an allocation that distributes the resource only between the individuals with powers 5 and 9, with $(5/14, 9/14)$. However it can only do so if $(5/14, 9/14)$ is a core allocation for the game consisting of these two individuals. It can be verified, however, that this is not the case, since in this game individual with power 9 is a winning majority and can allocate all of the resource to himself. By the same argument, no other coalition can implement a feasible allocation that gives to pay of greater than $(5/25, 9/25, 11/25)$ to its members. Therefore $(5/25, 9/25, 11/25)$ is in the core. It can also be verified that no other payoff vector is in the core, so that $(5/25, 9/25, 11/25)$ is the unique core allocation.

Our main result in this section is:

Theorem 5 *For any* $\Gamma_N = \left(N, \{\gamma_i\}_{i \in N}, \alpha, v_N(\cdot)\right)$ *with* v_N *defined by (4), the core* $C(\Gamma_N)$ *is nonempty. Moreover, suppose that Assumption 1 holds. Then the core* $C(\Gamma_N)$ *is a singleton given by* $\phi(N)$ *as defined in Theorem 1.*

Proof. See Appendix B. ■

This result shows the equivalence between the axiomatic approach in Section 3, the dynamic game in the previous section and the cooperative game in this section. 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. This is captured by two features: first, in addition to zero payoffs and status quo allocation, only payoff vectors that correspond to core allocations for a smaller game are feasible (recall Definition 11)); second, only winning coalitions can implement feasible pay of factors (recall Definition 12)). These two features introduce the power and enforcement constraints that also featured in the axiomatic and the dynamic game approaches. In view of this, the finding that the set of core allocations here correspond to the self-enforcing ruling coalitions is perhaps not surprising, though still reassuring.

6 The Structure of Ruling Coalitions

6.1 General Results

In this section, we present several results on the structure of self-enforcing ruling coalitions. Given the equivalence results in the previous two sections, without loss of any generality we focus on

self-enforcing ruling coalitions of abstract games $\Gamma = \left(N, \{\gamma_i\}_{i \in N}, \alpha\right)$. We start with the following lemma which will be useful for the results that will follow.

Lemma 1 *Consider a game* $\Gamma = \left(N \cup M, \{\gamma_i\}_{i \in N}, \alpha\right)$ *with arbitrary disjoint finite sets* M *and* N *and suppose that Assumption 1 holds. 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. ■

This lemma implies that there is some amount of continuity in the structure of self-enforcing ruling coalitions, in the sense that the addition of a set of agents with limited powers to the society does not change the self-enforcing winning coalition.

The next proposition answers some of the central questions related to the types of ruling coalitions that can emerge under majority rules. In particular, the first two parts establise that two-person coalitions cannot emerge as the self-enforcing ruling coalitions, but any other size coalition can emerge as the ruling coalition in a society of arbitrary size. This result implies that relatively little can be said about the structure of ruling coalitions without putting some more structure.

Also for future use, the third and the fourth parts of this proposition establish that self-enforcing coalitions are fragile in the sense that addition or subtraction of a single agent from these coalitions or a combination of two self-enforcing coalitions leads to a non-self enforcing coalition.

Proposition 2 *Consider a game* $\Gamma = \left(N, \{\gamma_i\}_{i \in N}, \alpha\right)$ *with* $\alpha = 1/2$ *and suppose that Assumption 1 holds. Then:*

1. *For any n and m such that $1 \leq m \leq n$, $m \neq 2$, there exists a set of players N, $|N| = n$, with powers $\{\gamma_i\}$ 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.*

3. *Suppose that two coalitions of N, X and Y are both self-enforcing. Then coalition $X \cup Y$ is not.*

4. *Suppose that 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)** Given Lemma 1, it is sufficient to show that there is a self-enforcing coalition M of size m (then adding $n - m$ players with negligible powers to form coalition N would yield $\phi(N) = \phi(M) = M$). Let $i \in \{1, \ldots, m\}$ 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 $k = 2, 3, \ldots, m-1$, $\gamma_m = \sum_{j=1}^{m-1} \gamma_j - 1$.

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)** This follows from Example 1 combined with Assumption 1.

(Part 3) 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 self-enforcing ruling coalition in $X \cup Y$.

(Part 4) For the case of adding, it follows directly from Part 3, since coalition of one person is always self-enforcing. For the case of deleting: suppose that it is wrong, and the coalition is self-enforcing. Then, by Part 3, adding this person back will result in an non-self-enforcing coalition. This is a contradiction which completes the proof of Part 4. ■

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

Proposition 3 *Consider a game* $\Gamma = \left(N, \{\gamma_i\}_{i\in N}, \alpha\right)$ *with* $\alpha > 1/2$. *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* $\{\gamma_i\}$ *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 2. 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. However, again in the case of majority rule, we can provide a range of additional results.

Proposition 4 *Consider a game* $\Gamma = \left(N, \{\gamma_i\}_{i\in N}, \alpha\right)$ *with* $\alpha = 1/2$. *Consider the infinite sequence*

$$\bar{\gamma} = \left\{\frac{2^k - 1}{2^k}\right\}_{k=1}^{\infty}, \tag{5}$$

and let $\{\bar{\gamma}_i\}_{i\in N}$ *be the sequence* $\bar{\gamma}$ *truncated at* $|N|$. *Consider the game* $\Gamma = \left(N, \{\bar{\gamma}_i\}_{i\in N}, \alpha\right)$, *then the unique equilibrium ruling coalition* X^e *in this game has size* m *where*
$m = \max\left\{z \in \mathbb{Z}_+ : z = 2^k - 1 \text{ for } k \in \mathbb{Z}_+ \text{ and } z \leq |N|\right\}$.

Proof. First note that given the sequence of powers in (5) whenever $|X| > |Y|$, we have that coalition X is stronger than coalition Y, given the sequence $\{\bar{\gamma}_i\}$.

The rest of the proof is by induction. The claim is trivially true for $|N| = 1$. Suppose that it is proved for all sizes smaller than $|N|$. If $|N| = 2^k - 1$ for some $k \in \mathbb{Z}_+$, then all smaller winning coalitions have sizes from 2^{k-1} to $2^k - 2$; by induction we know that they are non-self-enforcing. Therefore, the equilibrium ruling coalition is the grand coalition. If, however, $|N| \neq 2^k - 1$ for any $k \in \mathbb{Z}_+$, then the coalition of strongest (first) m players is both winning (it consists of at least half of players, and if it is exactly half, it is the strongest possible half) and self-enforcing (by induction). Therefore, the minimal winning self-enforcing coalition is not N, completing the proof of the induction step and thus the proof. ■

The next proposition shows that an increase in the power of an individual can remove him out of the ruling coalition.

Proposition 5 *Consider two games* $\Gamma = \left(N, \{\gamma_i\}_{i\in N}, \alpha\right)$ *and* $\Gamma' = \left(N, \{\gamma'_i\}_{i\in N}, \alpha\right)$ *such that* $\gamma_i = \gamma'_i$ *for all* $i \neq j$ *and* $\gamma_j < \gamma'_j$. *Then it is possible that* $j \in \phi_\Gamma(N)$ *and* $j \notin \phi_{\Gamma'}(N)$.

Proof. This follows from the examples in the next subsection. ■

In coalition $(3,4,5,10)$, the most powerful individual, 10, is not a part of any ruling coalition. The next proposition establishes sufficient conditions for the most powerful individual need not be part of the ultimate ruling coalition.

Proposition 6 *Let $\{\gamma_i\}_{i\in N}$ be an increasing sequence such that any its of truncation satisfies Assumption 1. Denote the most powerful player by n (i.e., $\gamma_n > \gamma_i$ for all $i \in N$, $i \neq n$). Consider the game $\Gamma = \left(N, \{\gamma_i\}_{i\in N}, \alpha\right)$. Suppose that the grand coalition is not self-enforcing. If $\gamma_n \in \left(\frac{\alpha}{1-\alpha}\sum_{i=2}^{n-1}\gamma_i, \frac{\alpha}{1-\alpha}\sum_{i=1}^{n-1}\gamma_i\right)$, then the most powerful individual, n, is not a part of the self-enforcing ruling coalition.*

Proof. Inequality $\gamma_n > \frac{\alpha}{1-\alpha}\sum_{i=2}^{n-1}\gamma_i$ implies that any coalition that includes n, but excludes even the weakest player will not be self-enforcing. The inequality $\gamma_n < \frac{\alpha}{1-\alpha}\sum_{i=2}^{n-1}\gamma_i$ implies that n is not a winning coalition by himself. Therefore, either N is self-enforcing or $\phi(N)$ does not include the strongest player. Since N is not self-enforcing by hypothesis, the conclusion follows. ■

The next question is whether the grand coalition itself could be a self-enforcing coalition. Let $[z]$ denote the integer part of z. Then we have the following result:

Proposition 7 *Consider a game $\Gamma = \left(N, \{\gamma_i\}_{i\in N}, \alpha\right)$ and suppose that Assumption 1 holds. Then, we have:*

1. *Let $\alpha = 1/2$ and suppose that for any two coalitions $X, Y \subset N$ 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. *Rank the strength of the players in ascending order, $\gamma_1, \gamma_2, \ldots \gamma_{|N|}$. Then a sufficient condition for $|X| > |Y| \implies \gamma_X > \gamma_Y$ is*
$$\sum_{j=1}^{|N|}\left|\frac{\gamma_j}{\gamma_1} - 1\right| < 1. \tag{6}$$

3. *Suppose $\alpha \in [1/2, 1)$ and suppose that for any two coalitions $X \subset Y \subset N$ such that $|X| \gtrless \alpha|Y|$ we have $\gamma_X \gtrless \alpha\gamma_Y$. Then $\phi(N) = N$ if and only if $|N| = k_m$ where $k_1 = 1$ and $k_m = \left[\frac{k_{m-1}}{\alpha}\right] + 1$ for $m > 1$.*

Proof. (Part 1) Let us check that the condition in Part 3 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 3 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 3.

(Part 2) 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

$$\sum_{j \in X'} \frac{\gamma_j}{\gamma_1} \leq \sum_{j \in Y'} \frac{\gamma_j}{\gamma_1}$$

and thus

$$\sum_{j \in X'} \left(\frac{\gamma_j}{\gamma_1} - 1 \right) + \left| X' \right| \leq \sum_{j \in Y'} \left(\frac{\gamma_j}{\gamma_1} - 1 \right) + \left| Y' \right|.$$

Rearranging, we have

$$1 \leq \left| X' \right| - \left| Y' \right| \leq \sum_{j \in Y'} \left(\frac{\gamma_j}{\gamma_1} - 1 \right) - \sum_{j \in X'} \left(\frac{\gamma_j}{\gamma_1} - 1 \right) \leq \sum_{j \in X' \cup Y'} \left| \frac{\gamma_j}{\gamma_1} - 1 \right|.$$

However, X' and Y' do not intersect, and therefore this violates (6). This contradiction completes the proof of Part 2.

(Part 3) The proof is by induction. The base is trivial: a one-player coalition is self-enforcing, and $|N| = k_1 = 1$. Now assume the claim has been proved for all $q < |N|$, let us prove it for $q = |N|$. If $|N| = k_m$ for some m, then any winning (within N) coalition X must have size at least $\alpha \left(\left[\frac{k_{m-1}}{\alpha} \right] + 1 \right) > \alpha \frac{k_{m-1}}{\alpha} = k_{m-1}$ (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. ■

Proposition 8 *Consider a game* $\Gamma = \left(N, \{\gamma_i\}_{i \in N}, \alpha \right)$ *and suppose that Assumption 1 holds. There exists* $\delta > 0$ *such that if* $\max_{i,j \in N} \left\{ \frac{\gamma_i}{\gamma_j} \right\} < 1 + \delta$*, then* $\phi(N) = N$ *if and only if* $|N| = k_m$*, where* $k_m = \left[\frac{k_{m-1}}{\alpha} \right] + 1$ *and* $k_0 = 1$.

Proof. The proof is again by induction. The base is trivial: a one-player coalition is self-enforcing, and $|N| = k_0 = 1$. Now assume this proposition has been proved for all $|N| < q$, let us prove it

for $|N| = q$. Take any distribution of powers $\{\gamma_i\}_{i\in N}$ satisfying Assumption 1 close to $(1, 1, \ldots, 1)$ in the sense that $\max_{i,j\in N}\left\{\frac{\gamma_i}{\gamma_j}\right\} < 1 + \delta$. We will now prove that if $|N| = k_m$ for some m, then any winning coalition must have size greater than $\alpha\left(\left[\frac{k_{m-1}}{\alpha}\right] + 1\right) > \alpha\frac{k_{m-1}}{\alpha} = k_{m-1}$. By induction, all such coalitions are not self-enforcing, and this means that the grand coalition is self-enforcing. The complete proof, consider the case where $|N| \neq k_m$ 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 also self-enforcing by induction (because it is close to the center $(1, 1, \ldots, 1)$ of size k_{m-1}). It is also winning; this follows because $k_{m-1} = \alpha\frac{k_{m-1}}{\alpha} \geq \alpha\left[\frac{k_{m-1}}{\alpha}\right] = \alpha(k_m - 1) \geq \alpha|N|$, 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 winning self-enforcing coalition which is not the grand coalition, which implies that the grand coalition is not self-enforcing. This completes the proof. ■

The next proposition establishes another continuity result that if powers are 'close' in two different games, then these two games will have the same self-enforcing ruling coalition. The state and prove this proposition, endow the set of sequences γ, $\mathbb{G}$, with the sup-metric, with distance given by $\rho(\gamma, \gamma') = \max_{i=1,\ldots,|N|}|\gamma_i - \gamma'_i|$. Define a δ-neighborhood of γ as $\{\gamma' \in \mathbb{G} : \rho(\gamma, \gamma') < \delta\}$.

Proposition 9 *Consider two games* $\Gamma = \left(N, \{\gamma_i\}_{i\in N}, \alpha\right)$ *and* $\Gamma' = \left(N, \{\gamma'_i\}_{i\in N}, \alpha\right)$. *There exists* $\delta > 0$ *such that if* $\gamma' = \{\gamma'_i\}_{i\in N}$ *lies within* δ*-neighborhood of* γ *then* $\phi_\Gamma(N) = \phi_{\Gamma'}(N)$.

Proof. This theorem follows from openness property in Theorem 6 in Appendix C. ■

6.2 Self-Enforcing Coalitions When $N = 3$

In this subection, we illustrate our basic findings about the structure of equilibrium ruling coalitions for $N = 3$. We also use this representation to show that even in this most simple environment, an increase in α might make it less likely that larger coalitions emerge as the self-enforcing 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 any generality). As discussed in the Introduction, there are two kind of constraints that define the set of all self-enforcing coalitions: "power constraints" $\left\{\sum_{j\in K}\gamma_j = \alpha\right\}$ which are always parallel to be respective $(K-1)$-dimensional edge, and "enforcement constraints," which correspond to cones.

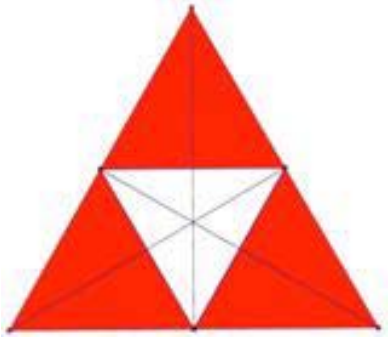

Figure 3: $\alpha = \frac{1}{2}$.

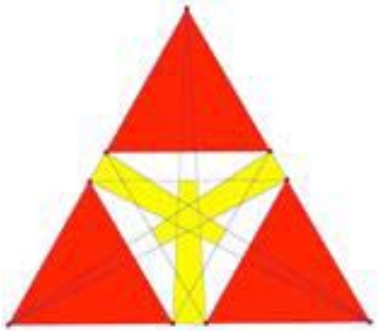

Figure 4: α between $\frac{1}{2}$ and $\frac{2}{3}$

We now show 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 powers such that the ground coalition is the self-enforcing ruling coalition is shown in white *white*. The *red* area is the set of configurations which are dominated by a single member. The yellow area is the set of points where a two-member coalition is both self-enforcing and winning.

Figure 3 corresponds to the case $\alpha = 1/2$. For any point $(\gamma_1, \gamma_2, \gamma_3)$ outside the white triangle, there is some member i who has power $\gamma_i > 1/2$. Figure 4 corresponds to the case when α becomes larger than 1/2; it demonstrate that set of self-enforcing coalition, though still a joint of a finite number of convex sets, has 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 non-self-enforcing. When α is large enough, but still less than 2/3, the self-enforcing coalition set does become a joint of a finite number of convex connected components (namely, of three trapezoids). When $\alpha = 2/3$, the trapezoids become triangles.

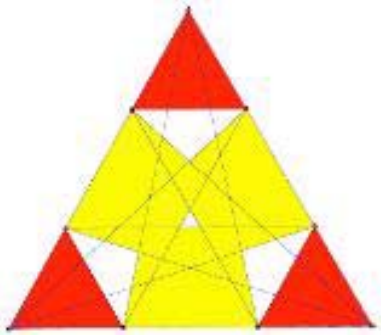

Figure 5: $\alpha > \frac{2}{3}$

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 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, $2/3 < \alpha \leq 1$, it is a joint of four triangles as on Figure 5. Obviously, points in the "middle" set of self-enforcing coalitions are even more "stable" than other self-enforcing coalitions: even if there is a random shock that eliminates some players, the remainder is self-enforcing coalition (in the respective game).

7 Extensions

7.1 Party Formation

If there were no enforcement constraints, the minimal winning coalition would always emerge as the ruling coalition. 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 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 a two-stage game where in the first stage, individuals form a bind-

ing agreement (a party), and then in the second period, they play the game above. The result will be that the minimal winning coalition will form a party (entering into a binding agreement) guaranteeing power for its members.

Proposition 10 *In the party-formation game, the unique ultimate ruling coalition X is the minimal winning coalition, i.e. X is such that $\sum_{i\in X}\gamma_i > \alpha\sum_{i\in N}\gamma_i$ and for any $Y \in P(N)$ with $\sum_{i\in Y}\gamma_i > \alpha\sum_{i\in N}\gamma_i$, $\sum_{i\in X}\gamma_i < \sum_{i\in Y}\gamma_i$.*

Proof. The proof follows the steps of the proof of Theorem 3 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). The next result investigates this question.

Proposition 11 *Suppose that $\alpha = 1/2$, Assumption 1 holds and X is a minimal winning coalition in N that is not self-enforcing. Suppose that under the unique self-enforcing coalition, each $j \in N$ receives w_j. Then for any $\eta > 0$, there exists a redistribution of power among the members of X such that for some $i \notin X$, $X \cup \{i\}$ becomes the self-enforcing ruling coalition, and implements a payoff $\hat{w}_j$ for all $j \in N$, such that $\hat{w}_j > w_j$ for all $j \in X$ and $\hat{w}_i < \eta$.*

Proof. Without loss of generality, assume that $\sum_{i\in N} w_i = 1$. If $|X| = 1$ or $|X| = 2$, the statement is trivial since the minimal winning coalition is always self-enforcing. Suppose that $|X| \geq 3$. Let k be the strongest member of X, i.e., $w_k \geq \max_{i\in X} w_i$. Consider the parametrized family $(w_i^\beta)_{i\in X}$ of distributions of power in coalition X : $w_k^\beta = w_k + \left(\sum_{i\in N\setminus X} w_i\right)\beta$ and $w_i^\beta = w_i + \left(\sum_{i\in N\setminus X} w_i\right)(1-\beta)$. When $\beta = 1$, k alone forms a winning coalition, since $w_k + \sum_{i\in N\setminus X} w_i \geq \sum_{i\in X\setminus\{k\}} w_i$. (Otherwise, coalition $X\setminus k$ is winning which contradicts the minimality of X.) We claim that there exists some β such that $\left(w_i^\beta\right)_{i\in X}$ is a self-enforcing coalition.

Let β_0 be determined by $w_k + \left(\sum_{i\in N\setminus X} w_i\right)\beta_0 = \sum_{i\in X\setminus\{k\}} w_i + \left(\sum_{i\in N\setminus X} w_i\right)(1-\beta_0)$. (Such β_0 exists since $w_k < \sum_{i\in N\setminus\{k\}} w_i$ by assumption). Since $0 < \beta_0 < 1$, $w_i^{\beta_0} > w_i$ for any $i \in X$. Thus, the remaining task is to ensure that XThere are two possibilities: either coalition $(w_i^{\beta_0})_{i\in X\setminus\{k\}}$ is not self-enforcing, or it is not.

Suppose that coalition $(w_i^{\beta_0})_{i\in X\setminus\{k\}}$ is not self-enforcing. We claim that there exists $\delta > 0$ such that coalition $\left(w_k^{\beta_0} - \delta, \left(w_i^{\beta_0} + \frac{\delta}{|X|-1}\right)_{i\in X\setminus\{k\}}\right)$ is self-enforcing (we leave player m with $w_m^{\beta_0} = 0$).

Take δ small enough so that the ordering of every two subcoalitions of the set $X\backslash\{k\}$ does not change. (we could do this because of Assumption 1). We could also assume that δ is small enough so that for any $j \in X\backslash\{k\}$, $w_k^{\beta_0} - \delta > \sum_{i\in X\backslash\{k,j\}} \left(w_i^{\beta_0} + \frac{\delta}{|X|-1}\right)$. This implies that $w_k^{\beta_0} - \delta$ cannot be a member of any proper self-enforcing subcoalition in X. We also have $w_k^{\beta_0} - \delta < \sum_{i\in X\backslash\{k\}} \left(w_i^{\beta_0} + \frac{\delta}{|X|-1}\right)$, since $w_k^{\beta_0} = \sum_{i\in X\backslash\{k\}} w_i^{\beta_0}$. Since k cannot be a member of any proper subcoalition of X, and cannot be eliminated by the rest, coalition X is self-enforcing. Now let $\hat{w}_k = w_k^{\beta_0} - \delta$, $\hat{w}_i = w_i^{\beta_0} + \frac{\delta}{|X|-1}$, $i \in X\backslash\{k\}$, and $\hat{w}_m = 0$ and the proof is complete.

Now suppose that coalition $(w_i^{\beta_0})_{i\in X}$ is self-enforcing. We claim that in this case coalition $\left(w_1^{\beta_0} - \frac{\delta}{3}, (w_i^{\beta_0} - \frac{\delta}{4(|X|-1)})_{i\in X\backslash\{k\}}, w_m^{\beta_0} = \frac{5}{12}\delta\right)$ is self-enforcing for any sufficiently small $\delta > 0$. Indeed, we have $w_k^{\beta_0} - \frac{\delta}{3} < \sum_{i\in X\backslash\{k\}} \left(w_i^{\beta_0} - \frac{\delta}{4(|X|-1)}\right) = \sum_{i\in X\backslash\{k\}} w_i^{\beta_0} - \frac{\delta}{4}$, and therefore k would not support eliminating m, while m clearly would not support eliminating either of players in X. Thus, $\left(w_1^{\beta_0} - \frac{\delta}{3}, (w_i^{\beta_0} - \frac{\delta}{4(|X|-1)})_{i\in X\backslash\{k\}}, w_m^{\beta_0} = \frac{5}{12}\delta\right)$ is self-enforcing. By construction, δ might be taken to satisfy $\delta < \eta$. Let $\hat{w}_k = w_k^{\beta_0} - \frac{\delta}{3}$, $\hat{w}_i = w_i^{\beta_0} - \frac{\delta}{4(|X|-1)}$, $i \in X\backslash\{k\}$, and $\hat{w}_m = \frac{5}{12}\delta$ and the proof is again complete. ■

The two following conjectures are natural and remain to be proved. First, Proposition 11 should hold for any $\alpha > 1/2$. Second, there exists a redistribution of power that makes X rather than $X \cup \{i\}$ self-enforcing.

8 Conclusion

The central question of political economy is how collective choices are made among a group of individuals with conflicting preferences. 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 three axioms: first, that the ruling coalition should be powerful enough (the *power constraint*); second, that the ruling coalition should be self-enforcing (the *enforcement constraint*); and third, that individuals should always choose the coalition that gives them higher returns (individual rationality). We show that there exists a unique mapping, which is also single-valued, that satisfies these three axioms. This provides an axiomatic way of characterizing the self-enforcing ruling coalitions for any game.

We support this notion by showing that the result of our axiomatic analysis also follows as "reasonable equilibria" of a dynamic game of coalition formation and also as the unique core allocation of a related non-transferable cooporative game. In particular, we construct a simple dynamic game that encompasses the same notions that a ruling coalition should have as certain amount of power and should be self-enforcing. 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, non-cooperative and cooperative games, we present a series of results on the structure of self-enforcing ruling coalitions. In particular, we show the following results:

1. There always exists a self-enforcing ruling coalition and can be computed by induction.

2. 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.

3. 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.

4. 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.

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

6. 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.

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 self-enforcing 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). The 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 talk of more or less "robust" coalitions and also 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: Examples

8.1 SPNE and MPE in the Main Game

Example 4 Let $n = 4$, $\alpha = 1/2$, $(\gamma_1, \gamma_2, \gamma_3, \gamma_4) = (4, 5, 6, 8)$. Let X be any 3-member subset of N. Define strategies as follows. Before any player is eliminated, members from X propose coalition X and the player from $N \setminus X$ proposes coalition N; if any coalition other than X is proposed, everyone votes 'against', while if X is proposed, members from X vote for it. After some coalition Y is eliminated, the remaining players play some MTHPE of the corresponding game. The equilibrium we described is subgame perfect: indeed, after elimination a subgame perfect equilibrium is played by definition of MTHPE. Before elimination, no player has an incentive to deviate. Indeed, by one-shot deviation principle we may consider one-shot deviations only. At the voting stage, nobody has an incentive to switch from voting 'against' to voting 'for' some coalition other than X, because he does not form a winning coalition by himself and thus will not affect the voting outcome. If some player deviates from voting for X to voting against it, he will either change nothing or will make N the URC. However, as established in the text„ if X is accepted, it will be the URC, and this is preferrable to any of its members than as compared to N. Therefore, such deviation is not payoff-improving. As for agenda-setters, any of them knows that any suggestion other than X will not be accepted. Hence, a member of X will either change nothing or make N the URC by his deviation; as before, this deviation will not make him better off. Finally, the agenda-setter who is not a part of X makes an offer which is not accepted in equilibrium, and if he offers something different then either it will not be accepted, or it will be X which will be accepted. Clearly, the only case where his deviation will have a non-trivial effect is where he is the last person to make a proposal and he deviates to proposing X. However, this changes the URC from N to X, which makes him worse off. We have proved that this is a SPNE. Strategies are Markovian (by definition of MTHPE and by construction), and thus it is a MPE. However, in our reasoning X may be different, as we have thus constructed as much as 6 different MPEs with different coalitions as the URC.

8.2 THPE vs. MTHPE

Example 5 Consider a game of three players with extensive form and payoffs as shown on Figure 6. 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

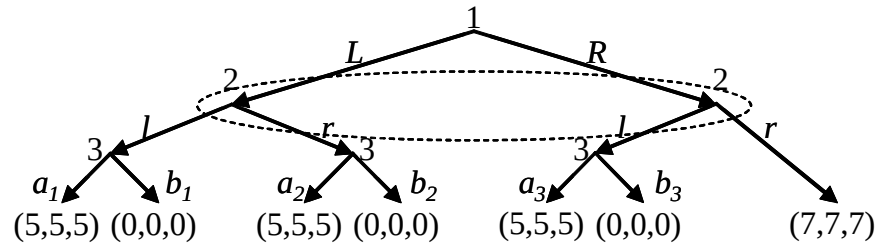

Figure 6: A Game With Herding in Trembling-Hand Perfect Equilibrium.

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 $u_L = 5\left(\left(1-\eta^3\right)\left(1-\eta^2\right) + \eta^3(1-\eta)\right) = 5 - 5\eta^2 - 5\eta^4 + 5\eta^5$. 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 5 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.

8.3 SWDE and MTHPE

Example 6 Consider a game of two players with extensive form depicted on Figure 7. 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.

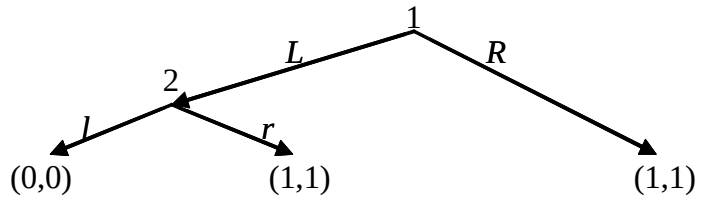

Figure 7: A Game With SWDE which is not MTHPE

Appendix B: Omitted Proofs

Proof of Theorem 2: (sketch) Let us first remember the notation in Proposition 1. In particular, let the order $\succ$ over sets be such that $X' \succ X''$ if and only if $\gamma_{X'} > \gamma_{X''}$ and the min operator over sets according to this order. For $k = 1, ..., |N|$, let $\mathbf{X}^k = X^k \cup \mathcal{S}\left(X^k\right)$, where $\mathcal{S}\left(X^k\right) \equiv \left\{Y \in P\left(X^k\right) : \gamma_Y > \alpha\gamma_{X^k} \text{ and } \phi\left(Y\right) = Y\right\}$. Then the operator $\phi\left(N\right)$ satisfies

$$\phi\left(N\right) = \min_{X \in \mathcal{S}(N) \cup \{N\}} X.$$

Define also the set of self-enforcing coalitions that are winning within N and include individual i as:

$$\mathcal{S}_i\left(N\right) \equiv \mathcal{S}\left(N\right) \cap \left\{X \in P\left(N\right) \mid i \in X\right\}$$

and thus

$$\phi_i\left(N\right) = \min_{X \in \mathcal{S}_i(N) \cup \{N\}} X.$$

Now for any coalition $X \in P\left(X\right)$ define

$$\phi_X\left(N\right) = \min_{X \in \left(\bigcup\limits_{i \in X} \mathcal{S}_i(N)\right) \cup \{N\}} X \equiv \min_{i \in X} \phi_i\left(N\right).$$

Evidently, $\phi_N\left(N\right) \equiv \phi\left(N\right)$. For convenience, let $\phi_\varnothing\left(N\right) = N$, which clearly agrees with our previous definition of $\phi_X\left(N\right)$. We start with the following lemma, which establishes the most important part of the proof of the theorem.

Lemma 2 *Suppose the game has reached node ξ_{j_q} where there have been j $(0 \leq j \leq |N| - 1)$ eliminations, the first q $(0 \leq q \leq |N_j| - 1)$ proposals have been made by players $i_{j_1}, \ldots i_{j_q}$ were rejected, and the Nature as picked player $i_{j_{q+1}}$ as the next proposer. Let $Z_{j_q} \equiv N_j \setminus \left\{i_{j_1}, \ldots i_{j_q}\right\}$. Then*

1. *In any SWDE starting from node ξ_{j_q}, $\phi_{Z_{j_q}}\left(N_j\right)$ will be the URC. Moreover, the total number of eliminations will be j if $\phi_{Z_{j_q}}\left(N_j\right) = N_j$ and $j + 1$ otherwise.*

2. There exists a pure strategy SWDE where

(a) player $i_{j_{q+1}}$ makes proposal $X_{j_{q+1}} = \phi_{i_{j_{q+1}}}(N_j)$ and

(b) following the proposal $X_{j_{q+1}}$ by player $i_{j_{q+1}}$, player $i \in X_{j_{q+1}}$ votes for this proposal if $w_i\left(\phi\left(X_{j_{q+1}}\right)\right) > w_i\left(\phi_{Z_{j_{q+1}}}(N_j)\right)$ (where $w_i(X) = \gamma_i/\gamma_X$ if $i \in X$ and $w_i(X) = 0$ if $i \notin X$) and against it otherwise, so that proposal $X_{j_{q+1}}$ is accepted if and only if $\phi\left(X_{j_{q+1}}\right) = \phi(N)$ and $Z_{j_{q+1}} \cap \phi(N) = \varnothing$.

Proof. The proof follows uses double (backward) induction, on j and q. The base for outer induction is $j = |N| - 1$. If there were $|N| - 1$ eliminations, then, obviously, $|N_j| = 1$ and the only possible value for q is $q = 0$. Any play that starts from that node will end up with N_j as the URC, and the payoffs that all players will get are the same. It is therefore a pure strategy SWDE for the only player $i_{j_{q+1}}$ to offer $X_{j_{q+1}} = \phi_{i_{j_{q+1}}}(N_j) = N_j$ (actually, this is the only proposal he can make) and then to vote against it (because $w_i(N_j) = w_i\left(\phi_{\varnothing}(N_j)\right)$). Here, $\phi_{Z_{j_q}}(N_j) = \phi_{i_{j_{q+1}}}(N_j) = N_j$ as well, and thus Part 1 follows (again, no other URC is possible in this subgame). The total number of eliminations is $j = |N| - 1$, and this completes the proof of the base of the induction by j.

Now let us prove the step of the induction. Suppose that the claim in the lemma has been proved for all $j' > j$, let us prove it for $j' = j$. To do this, we use induction by q. The base is $q = |N_j| - 1$, so suppose that $i_{j_{q+1}}$ is the last player to make a proposal in the jth round. If his proposal $X_{j_{q+1}}$ (whatever it is) is rejected, then $\phi_{Z_{j_{q+1}}}(N_j) = \phi_{\varnothing}(N_j) = N_j$ becomes the URC. If this proposal is accepted, then (unless $X_{j_{q+1}} = N_j$, in which case the game ends with N_j as the URC) $(j+1)$th elimination takes place. By induction, the resulting URC will be $\phi\left(X_{j_{q+1}}\right)$ (note that this does not depend on the player $i_{(j+1)_1}$ that the Nature picks as the first proposer in the next round. Then, voting for the proposal if and only if $w_i\left(\phi\left(X_{j_{q+1}}\right)\right) > w_i\left(\phi_{Z_{j_{q+1}}}(N_j)\right)$ is a weakly dominant strategy for player $i \in X_{j_{q+1}}$ (holding future plays fixed). In particular, only proposals such that $\phi\left(X_{j_{q+1}}\right)$ is a winning coalition other than N_j are accepted; by definition, $\phi\left(X_{j_{q+1}}\right)$ is self-enforcing. Therefore, it is weakly dominant for player $i_{j_{q+1}}$ to make proposal $X_{j_{q+1}} = \phi_{i_{j_{q+1}}}(N_j)$. Indeed, if his proposal is rejected, he will get $w_{i_{j_{q+1}}}(N_j)$, while if it is accepted he will get $w_{i_{j_{q+1}}}\left(\phi\left(X_{j_{q+1}}\right)\right)$. As we have just shown, for a proposal X to be accepted in the subsequent voting it is necessary and sufficient that $\phi(X)$ is a winning self-enforcing coalition (other than N_j). If $i_{j_{q+1}}$ is not part of any such coalition (so that $\phi_{i_{j_{q+1}}}(N_j) = N_j$), then he is weakly better off from offering $\phi_{i_{j_{q+1}}}(N_j) = N_j$: indeed, any other offer $X \neq N_j$ is either rejected (so N_j becomes the URC) or accepted, in which case $\phi(X)$ will be the URC, and this would make

player $i_{j_{q+1}}$ worse off because he is not part of $\phi(X)$. The opposite case is where $\phi_{i_{j_{q+1}}}(N_j) \neq N_j$. Player $i_{j_{q+1}}$ knows that whatever proposal X he makes, if it accepted, then equilibrium future play will lead to a self-enforcing coalition $\phi(X)$, which is also winning within N_j (otherwise proposal X will not get the majority on the voting). Of all winning self-enforcing coalitions, player $i_{j_{q+1}}$ prefers $\phi_{i_{j_{q+1}}}(N_j)$. He can achieve this by proposing $X_{i_{j_{q+1}}} = \phi_{i_{j_{q+1}}}(N_j)$; moreover, in this case, only one extra $-\varepsilon$ will be subtracted from his utility. Obviously, he cannot do better than that, and hence proposing $X_{i_{j_{q+1}}} = \phi_{i_{j_{q+1}}}(N_j)$ is a weakly dominant action. The final step is to check Part 2 of the lemma. Take any SWDE. By induction, we know what would happen if player $i_{j_{q+1}}$'s proposal is accepted or rejected. If his proposal is some winning coalition $X_{i_{j_{q+1}}} \neq N_j$, players from $\phi\left(X_{i_{j_{q+1}}}\right)$ will vote for it, while those from $X_{i_{j_{q+1}}} \setminus \phi\left(X_{i_{j_{q+1}}}\right)$ will vote against it in this SWDE. Hence, any proposal $X_{i_{j_{q+1}}} \neq N_j$ such that $\phi\left(X_{i_{j_{q+1}}}\right)$ is a winning coalition within N_j will be accepted; other proposals (other than N_j) will be rejected in this SWDE. Given this, player $i_{j_{q+1}}$ maximizes his utility if he proposes $X_{i_{j_{q+1}}} = \phi_{i_{j_{q+1}}}(N_j)$. This completes the proof of the base of the inner induction.

Induction by q follows in a similar way and is omitted. To inductions together established the lemma. ■

Now we return to the proof of Theorem. The lemma is true for $j = 0$ and $q = 0$; moreover, the URC and the number of eliminations does not depend on the player i_{0_1} that the Nature picks as the first proposer. Since $\phi_{N_0}(N_0) = \phi(N)$, the conclusions in the Theorem follow. ■

Proof of Theorem 4:

(Part 1) This is proved by backward induction by 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 2) Consider a one-stage game with two players making simultaneous moves with payoff matrix

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

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 3) This follows from Example 6 in Appendix A. ■

Proof of Theorem 5: (sketch) Fix an abstract game $\Gamma = \left(N, \{\gamma_i\}_{i\in N}, \alpha\right)$ and construct the corresponding non-transferable utility cooperative game $\Gamma_N = \left(N, \{\gamma_i\}_{i\in N}, \alpha, v_N(\cdot)\right)$. Suppose that for the mapping ϕ defined in Theorem 1, $\phi(N) = X$ is uniquely determined (in view of Assumption 1). We first show that the allocation x with $w_i(X) = \gamma_i/\gamma_X$ for $i \in 0$ and $w_i(X) = 0$ for $i \notin X$ is in the set $v_N(X)$. Since $\phi(N) = X$, X must be winning within N according to Definition 1, thus $\gamma_X > \alpha\gamma_N$. Therefore, coalition X can implement any feasible allocation. Moreover, again since $\phi(N) = X$, X is self-enforcing. Therefore, in the cooperative game $\Gamma_X = \left(X, \{\gamma_i\}_{i\in X}, \alpha, v_X(\cdot)\right)$, the allocation x is in the core. Therefore, coalition X can implement x. By

definition, again since $\phi(N) = X$, there is no winning self-enforcing subcoalition of X, thus there exists no Y that can implement a feasible payoff vector that will give higher utility to its members than x. This establishes that x is in the core.

To see that there are no other allocation in the core, suppose that there exists some feasible allocation y in the core implemented by a coalition Y. Since both X and Y are winning, we must have $X \cap Y \neq \varnothing$. By construction, all i in $X \backslash X \cap Y$ have higher payoff under x than y. Again by definition of $\phi(N) = X$ and given Assumption 1, all i in $X \cap Y$ also have higher payoff under x than y. Moreover, by the same argument as the previous paragraph x he's a feasible a location that coalition X can implement. Therefore, coalition x can block y, and thus y is not in the core, completing the proof. ■

9 Appendix C: The Structure of Self-Enforcing Coalitions

In this appendix, we generalize the results on the structure of self-enforcing coalitions to situations in Assumption 1 does not hold. Let us define $\mathcal{S}(N)$ as the set of all winning self-enforcing coalitions. Formally, $\mathcal{S} : P(N) \to P(N)$ is a mapping that satisfies Axioms 1 and 2 in Section 3. Recall also that under Assumption 1 all elements of $\mathcal{S}(N)$ have different powers, and hence the winning self-enforcing coalition with minimal power, $\phi(N)$, is unique as proved in Theorem 1. Furthermore, denote the set of coalitions that satisfy the requirement in Assumption 1, i.e., that are generic, by $\mathcal{G}(N)$, and the set of coalitions that are generic but not self-enforcing by $\mathcal{N}(N)$.

Define the joint set of generic and non-generic self-enforcing coalitions by $\mathcal{S}^+(N)$i.e.,

$$\mathcal{S}^+(N) = \mathbb{R}^n_+ \setminus \overline{\left(\mathbb{R}^n_+ \setminus \overline{\mathcal{S}(N)}\right)},$$

where $\overline{A}$ denotes the closure of set A in the standard topology.

We then have:

Theorem 6 *1. The set of generic coalitions $\mathcal{G}(N)$, the set of self-enforcing coalitions $\mathcal{S}(N)$, and the set of coalitions which are generic but not self-enforcing $\mathcal{N}(N)$, are open sets in $\mathbb{R}^n_+$. The set $\mathcal{G}(N)$ is also dense in $\mathbb{R}^n_+$*

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

Proof. **(Part 1)** The set $\mathcal{G}(N)$ 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 $X \neq Y$. These hyperplanes are closed sets, hence, a small perturbation of powers of a generic point will leave preserve this

property (genericity). This ensures that $\mathcal{G}(N)$ is an open set; it is dense because hyperplanes have dimension lower than n. The proofs for $\mathcal{S}(N)$ and $\mathcal{N}(N)$ are by induction. The base is trivial: indeed, $\mathcal{S}(N) = \mathbb{R}_+$ and $\mathcal{N}(N) = \varnothing$ are open sets. Now suppose that we have proved this Proposition for all $n < k$. Take any generic coalition N with powers $\{\gamma_i\}$; it 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 $\{\gamma_i'\}$. If this perturbation is small, then the set of winning coalitions is the same, and, by induction, the set of proper self-enforcing coalitions is the same as well. Therefore, the perturbed coalition $\{\gamma_i'\}$ is self-enforcing if and only if the initial coalition with powers $\{\gamma_i\}$ is self-enforcing; which completes the induction step.

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

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.

Aumann, R. and R. 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, F. (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, M. (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, J. and S. 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.

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

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.

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.

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.

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

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

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.

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