

Dynamics and Stability of Constitutions, Coalitions, and Clubs*

Daron Acemoglu
MIT

Georgy Egorov
Harvard

Konstantin Sonin
New Economic School

April 2008

Abstract

A central feature of collective decision-making in many social groups, such as political coalitions, international unions, or private clubs, is that the rules that govern the procedures for future decision-making and the inclusion and exclusion of members are made by the current members and under the current regulations. This feature implies that dynamic collective decisions must recognize the implications of current decisions on future choices. For example, current constitutional change must take into account how the new constitution may pave the way for further changes in laws and regulations. We develop a general framework for the analysis of this class of problems. We provide both an axiomatic and a non-cooperative characterization of dynamically stable states and show that, under reasonable assumptions, these exist and are unique. We then apply our framework to a variety of problems in political economy, in coalition formation, and in the analysis of the dynamics of clubs. Major insights that emerge from this framework are: (1) A particular social arrangement is made stable by the instability of alternative arrangements that are preferred by sufficiently many members of the society. For example, stability of a constitution does not require absence of powerful groups opposing it, but the absence of an alternative constitution favored by powerful groups. (2) Efficiency-enhancing changes are often resisted because of further social changes that they will engender. Consequently, Pareto inefficient social arrangements often emerge as stable outcomes.

Keywords: club theory, constitutions, dynamic coalition formation, political economy, stability.

JEL Classification: D71, D74, C71.

Work in Progress. Comments Welcome.

*Daron Acemoglu gratefully acknowledges financial support from the National Science Foundation. We thank Robert Powell and participants of ASSA 2008 conference (New Orleans) and a seminar at Harvard University for helpful comments.

1 Introduction

Consider the problem of a society choosing its constitution, which will determine economic payoffs and the procedures for future decision-making. The current rewards from adopting a specific constitution will naturally be important in this decision. But, as long as the members of the society are forward-looking and patient, the future implications of the constitution may be even more important. For example, a constitution that encourages economic activity and benefits the majority of the population may nonetheless create future instability or leave room for a minority to seize control of the decision-making process. If so, the society—or the majority of its members—may rationally shy away from adopting such a constitution.

Many problems in political economy, club theory, organizational economics, and industrial organization have a structure resembling this example of constitutional choice. Consider, for example, the problem of a club choosing its membership, recognizing that new members will take part in the future expansion or contraction decisions. Another example is the problem of dynamic coalition formation, where parties forming a coalition recognize that members of the coalition will have a say over policy choices and the survival prospects of the coalition. Similar considerations arise in the context of an organization deciding how to restructure itself and how much power to give to a CEO or to a board of directors.

In this paper, we develop a general framework for the analysis of dynamic group-decision-making over constitutions, coalitions, and clubs. Although our model is motivated by political economy applications, it is general enough to nest the examples mentioned above (and a range of related problems discussed in the literature; see below). Formally, we consider a society consisting of a finite number of infinitely-lived individuals. The society starts in a particular *state*, which can be thought of as the *constitution* of the society, regulating how economic decisions are made. This state determines stage payoffs and also how the society can determine its future states (constitution), for example, which subsets of individuals can decide to reform the constitution and which other subsets can block such decisions, etc. This is a dynamic game of nontransferable utility (since the current constitution determines the current payoff for each member of the society). Our focus is on (Markov perfect) equilibria of this dynamic game when individuals are sufficiently forward-looking. In particular, we show the existence and characterize the structure of (*dynamically*) *stable states*, which are defined as states that arise and persist (repeat themselves).

Our analysis has two parts. The first part focuses on an axiomatic characterization of stable states. This part is motivated by the observation that when individuals are sufficiently forward-

looking, an individual will not wish to support change towards a state (constitution) that might ultimately lead to another, less preferred state. This notion can be captured by a simple *stability axiom*. Using this stability axiom, together with two other natural axioms and a set of minimal assumptions on the acyclicity of preferences, we characterize the set of “outcome mappings”. These mappings determine all stable states and we show that each equilibrium outcome mapping defines exactly the same set of stable states. Under an additional assumption (on pairwise comparability of alternative desirable states), we also show that this mapping determines a unique dynamically stable state. This axiomatic characterization is simple and captures the economic essence of the examples mentioned above and of those that will be discussed in greater detail below.

The second part of our analysis provides a natural extensive-form noncooperative game capturing the same economic forces as those emphasized in the axiomatic characterization. It then shows that, under the same assumptions as above, and provided that the discount factor of the individuals is greater than a threshold, there exists an (essentially) unique pure strategy Markov Perfect Equilibrium, in the sense that all equilibria lead to the same unique dynamically stable state. Moreover, this dynamically stable state is identical to the stable state characterized in the axiomatic analysis.¹

An attractive feature of this analysis is that the set of dynamically stable states can be characterized recursively. This characterization is not only simple (the set of dynamically stable states can be computed using induction), but it also emphasizes a fundamental insight: a particular state is dynamically stable only if there does *not* exist another state that is dynamically stable and is preferred by a set of players that form a winning coalition within the current state.

Both the axiomatic and the noncooperative approaches emphasize the same economic forces, in particular, the natural lack of *commitment* that exists in dynamic decision-making—those that gain additional decision-making power as a result of reform (change in constitution or expansion in club size etc) cannot commit to refraining from certain further choices that would hurt the initial set of decision-makers. This lack of commitment is at the root of the most important general results of our framework, which can be summarized as follows:

1. A particular social arrangement (constitution, coalition or club) is made stable not by the absence of a powerful set of players that prefer another alternative, but because of the absence of an alternative *stable* arrangement that is preferred by a sufficiently powerful

¹An additional assumption in our noncooperative analysis is that there is a transaction cost incurred by all individuals every time there is a change in the state. This assumption is used to prove the existence of a pure strategy equilibrium and to rule out cycles without imposing stronger assumptions on preferences.

constituency. This insight suggests that to understand why certain institutional or other social outcomes persist, we need to understand the instabilities that changes away from these arrangements would unleash. It also highlights why in such dynamic environments, “minimum winning coalitions” are not relevant (because they themselves may be unstable).

2. Dynamically stable states will often be inefficient—in the sense that they may be Pareto dominated by the payoffs in another state. A transition to the Pareto dominating state would not be undertaken because this latter state may not be stable.

We next provide a brief outline of some of the applications that illustrate these general results. We also use these examples to discuss the related literature.

Example 1 (Inefficient Inertia) As an introductory example, consider a society consisting of two individuals (or two social groups), E and M . E represents the elite and initially holds power, whereas M corresponds to the middle class. There are three states. The first is absolutist monarchy, a , in which E rules, with no checks and there are no political rights for M . The second is constitutional monarchy, c , in which E rules but with checks and balances, so that M has greater security and is willing to invest. The final state is democracy, d , where M becomes more influential and the privileges of E disappear. Reflecting this situation, suppose that stage payoffs satisfy

$$w_d(E) < w_a(E) < w_c(E),$$

and

$$w_a(M) < w_c(M) < w_d(M).$$

In particular, $w_a(E) < w_c(E)$ means that E has higher payoff under constitutional monarchy than under absolutist monarchy, for example, because greater investments by M increase tax revenues. M clearly prefers democracy to constitutional monarchy and is least well-off under absolutist monarchy. Both parties discount the stage payoffs with discount factor $\beta \in (0, 1)$. As described above, “states” not only determine payoffs but also specify decision rules. In absolutist monarchy, E decides which regime will prevail tomorrow. In constitutional monarchy, however, M accumulates enough wealth to contest E ’s political power. To simplify the discussion in this example, suppose that starting in both regimes c and d , M decides next period’s regime. Now, starting in regime a , E must choose whether to stay in a or to undertake a reform towards c (clearly a move to d is not desirable). However, E understands that once the state becomes c , M will become sufficiently powerful and implement another reform towards d . Therefore, the

choices facing E are between staying in a forever and undertaking a reform to c , which will then lead to d , giving E continuation utility

$$\frac{w_d(E)}{1-\beta} < \frac{w_a(E)}{1-\beta}.$$

It is straightforward to see that with β sufficiently large, the unique dynamically stable state starting with a is a . Moreover, this example also illustrates the potential inefficiency of dynamically stable states. Both E and M would be strictly better off in state c rather than in the dynamically stable state a . This example therefore illustrates both general messages from our analysis. First, a particular state, here state a , is made stable by the absence of another state that is preferred by those who are powerful in the current state. In particular, in absolutist monarchy, the elite, E , is politically powerful and the alternative state preferred by the elite, c , is not stable. Second, Pareto inefficiency can emerge easily. In this example, both the elite and the middle class are better off in c than in a , but a persists. Thus the dynamically stable state is Pareto inefficient. If we add a third group without any political power, but which prefers a to c , the equilibrium would no longer be Pareto inefficient, but it would continue to be winning coalition inefficient because the winning coalition in state a , $\{E\}$, is better off in state c .

The same reasoning would apply if there were more states in-between c and d . For example, the first reform to c could lead to c' , then to c'' , then to c''' , and so on, until the chain ends up at d . The states in-between need not be different political regimes. They could correspond to different laws or regulations within the same political regime (and they may also all provide greater utility to E than state a). Nevertheless, the same reasoning implies that as long as β is sufficiently close to 1 and $w_d(E) < w_a(E)$, E would not initiate the reform.

A similar game can also be used to represent the implications of concessions in wars. For example, a concession that will increase the payoffs to both warring parties may not take place because it will change the future balance of power. Relatedly, a country may prefer to declare war that is costly both to itself and to its opponent rather than allow its opponent to become stronger over time and demand concessions in the future.

Finally, this example could also be used to illustrate how organizations might act “conservatively” and resist efficiency-enhancing restructuring. For instance, the appointment of a CEO who would increase the value of the firm may not be favored by the board of directors if they forecast that, down the line, the CEO would become powerful and reduce their privileges.²

²Ideas related to this example have been discussed in a number of different (but connected) contexts. Robinson (1997) and Bourguignon and Verdier (2000) discuss how a dictator or an oligarchy may refrain from providing productive public goods or from educational investments, because they may be afraid of losing power. Rajan and Zingales (2000) also emphasize similar ideas and apply them in the context of organizations. Acemoglu

Example 2 (Voting in Clubs) A richer environment also covered by our framework is the problem of voting in clubs considered first in the seminal-unpublished paper by Roberts (1999). Consider a society consisting of N individuals. Any subset of these individuals can become a club. If the current club is X_t , then each individual receives a stage payoff $w_i(X_t)$, and current club members decide (according to some voting rule, which might be weighted or unweighted, majority or supermajority voting) whether the club should contract or expand, that is, they decide what tomorrow’s club, X_{t+1} , should be. Roberts studies a special case of this environment. In particular, he assumes that individuals are ordered, $j = 1, 2, \dots, N$, and X_t must be an ordered subset of the form $\{1, \dots, J\}$ for some $J = 1, 2, \dots, N$. Finally, Roberts assumes that individual preferences satisfy the following “single-crossing” property: if individual j (weakly) prefers $\{1, \dots, J\}$ to $\{1, \dots, J - 1\}$, then all $j' > j$ have the same (strict) preference. Under these assumptions, Roberts establishes the existence of a mixed strategy Markov Voting Equilibrium (where a transition happens unless it is blocked by at least half of club members, see subsection 6.3) and of a Median Voter Equilibrium (where the most preferred choice of the median voter is implemented), and characterizes some of their properties. Our model nests a considerably more general version of this environment and enables us to establish the existence of a unique dynamically stable club (and a pure strategy equilibrium). In addition, it provides a more complete characterization of these dynamically stable states under weaker assumptions. Our analysis further clarifies why majorities will not necessarily opt to move towards club sizes that will increase their stage payoffs and why the resulting dynamically stable state (club) may be inefficient—both of these are related to the natural commitment problems inherent in dynamic collective decision-making mentioned above.

Barbera, Maschler, and Shalev (2001) is another important paper with similar insights. They also recognize that decisions on club size should take into account changes in the identity of pivotal agents. They consider a game in which any member of the club might unilaterally admit a new agent and characterize the equilibria of this dynamic game. Alesina, Angeloni, and Etro (2005) apply a simplified version of Roberts’s model to study the question of how current members should approach problems related to the expansion of the European Union. Alesina, Angeloni, and Etro’s model is also a special case of our general setup.

Finally, a modified version of this game can also be used to analyze franchise extension, as in Acemoglu and Robinson (2000, 2006a), Lizzeri and Persico (2004), or Jack and Lagunoff (2006).

and Robinson (2006a) construct a dynamic model in which the elite may block technological improvements or institutional reforms, because they will destabilize the existing regime. Fearon (1996, 2004) and Powell (1998) discuss similar ideas in the context of civil wars and international wars, respectively.

Example 3 (Self-Stable Constitutions) Another interesting and important political economy question related to the ideas discussed here is considered in Barbera and Jackson (2004). In the first period of the game, all individuals are identical and choose a constitution. In the second and final period, individuals receive information about their ultimate preferences regarding the comparison of a status quo and an alternative. Barbera and Jackson characterize “self-stable” constitutions (which may have a different supermajority rule for reforming the constitution itself) that will remain in place after additional information arrives. Aidt and Giovanni (2004) and Messner and Polborn (2004) also consider related problems, where constitutions impose different supermajority requirements for decisions on different issues. As we will discuss in Section 6 an infinite-horizon version of this class of games is also a special case of our general environment.

Example 4 (Dynamic Coalition Formation in Nondemocracies) In Acemoglu, Egorov, and Sonin (2008), we considered the problem of dynamic coalition formation in nondemocratic societies, where subsets of a ruling coalition cannot commit to not sidelining the remaining members. Payoffs are realized only after the game ends. A more general version of this environment, where coalition members receive stage payoffs at each date and new members can be brought in to be part of the ruling coalition, will also be shown to be a special case of the environment studied here.

Examples 1-4 illustrate a subset of economic and political problems that can be analyzed as special cases of our model. We view the rich set of environments that are covered by our model and the relative simplicity of the resulting dynamic stable states as major advantages of our approach. We believe that both the specific results derived here and the general ideas can be applied to a range of problems in political economy, organizational economics, club theory and other areas. Some of these additional examples are discussed in Section 6.

On the theoretical side, Roberts (1999) and Barbera, Maschler, and Shalev (2001) can be viewed as the most important precursors to our paper. An interesting and ambitious recent paper by Lagunoff (2006) constructs a general model of political reform motivated in part by Roberts (1999) and Barbera, Maschler, and Shalev (2001) as well as Acemoglu and Robinson’s (2000, 2006a) and Lizzeri and Persico’s (2004) analyses of franchise extension. Lagunoff’s (2006) approach is different from ours, since he focuses on “social choice rules” that represent different institutional environments and investigates whether a social choice rule will select itself. Lagunoff’s analysis provides an insightful condition in terms of “time inconsistency” of a social choice rule that determines whether a particular set of institutions will persist, though he

provides neither a uniqueness result nor a general characterization of the set of equilibria.³

The two papers most closely related to our work are Chwe (1994) and Gomes and Jehiel (2005). Chwe studies a model where payoffs are determined by states and there are exogenous rules governing transitions from one state to another. Chwe considers concepts of consistent sets and stable sets and reveals interesting links between between the two concepts. In Chwe's setup, it is not possible to obtain a noncooperative analysis or uniqueness or characterization results, while such results are at the heart of our paper. In particular, Chwe assumes that different transitions from the same state may require different coalitions, whereas in our paper a winning coalition can enforce any transition. Our assumption implies that a group that is powerful enough to block one transition is also powerful enough to block any other transition, which is a plausible and widely-applicable assumption. Gomes and Jehiel study a related environment where states determine payoffs and potential transitions, but focus on transferable utilities (that is, they allow side payments). Gomes and Jehiel show that a player may sacrifice his instantaneous payoff to improve his bargaining position for the future, which is related to the unwillingness of winning coalitions to transition to non-stable states in our paper. Gomes and Jehiel also establish potential inefficiencies, but for small discount factors. In contrast, in our game Pareto dominated outcomes are not only possible in general, but they typically emerge as unique equilibria and are more likely when discount factors are arbitrarily high. More generally, we also provide a full set of characterization (and uniqueness) results, which are not present in Gomes and Jehiel (and in fact, with side payments, we suspect that such results are not possible). Finally, in our paper a dynamically stable state depends on the initial state, while in Gomes and Jehiel, as the discount factor tends to 1, there is "ergodicity" in the sense that the ultimate distribution of states does not depend on the initial state.

Finally, our work is also related to the literature on club theory (see, for example, Buchanan, 1956, Ellickson et al., 1999, Scotchmer, 2001). While the early work in this area was static, a number of recent papers have investigated the dynamics of club formation. In addition to Roberts (1999) and Barbera, Maschler, and Shalev (2001), which were discussed above, some of the important papers in this area include Burkart and Wallner (2000), who develop an incomplete contracts theory of club enlargement, Jehiel and Scotchmer (2001), who show that the requirement of a majority consent for admission to a jurisdiction may not be more restrictive than an unrestricted right to migrate, and Bordignon and Brusco (2003). This last paper studies club enlargement in a dynamic environment and derives insights about the "enhanced

³See also Ray (2008) for a review of the literature on coalition formation and Baron (1991) on coalition formation in legislatures.

cooperation agreements,” which allow for sub-union formation inside the EU. Bordignon and Brusco show that, if the incumbent members can commit to a coordinated policy towards new members, the sub-union formation process can be efficient.

The rest of the paper is organized as follows. In Section 2, we introduce the general environment. Section 3 motivates and presents our axiomatic analysis. Section 4 introduces the dynamic noncooperative game and shows the equivalence between the Markov Perfect Equilibria of this game and the axiomatic solution of Section 3. Section 5 generalizes the baseline environment by allowing state-specific restrictions on transitions (i.e., not all reforms are possible starting in all states). Section 6 returns to the applications discussed above and provides a more detailed analysis of these applications, demonstrating how they can be treated as special cases of our model. Section 7 concludes, Appendix A contains the proofs omitted from the text and Appendix B provides a number of examples illustrating the roles of the assumptions used in the text.

2 Environment

In this section, we introduce the general environment.

There is a finite set of players \mathcal{I} . Time is discrete and infinite, indexed by t ($t \geq 1$). There is a finite set of *states* which we denote by \mathcal{S} . We denote the number of elements of the sets \mathcal{I} and \mathcal{S} by $|\mathcal{I}|$ and $|\mathcal{S}|$, respectively. Recall that these states may represent different institutions simply affecting payoffs, or constitutions that may affect both payoffs and the procedures for decision-making (e.g., the ruling coalition in power, the degree of supermajority, the weights or powers of different agents). Although our game is one of non-transferable utility, a limited amount of transfers can also be incorporated, by allowing multiple (but a finite set of) states that have the same procedure for decision-making, but reallocate payoffs across players.

The initial state is denoted by $s_0 \in \mathcal{S}$. This state can be thought of as being determined as part of the description of the game or as chosen by Nature according to a given probability distribution. For any $t \geq 1$, the state $s_t \in \mathcal{S}$ is endogenously determined. A non-empty set $X \subset \mathcal{I}$ is called *coalition*, and we denote the set of coalitions by \mathcal{C} (that is, \mathcal{C} is the set of nonempty subsets of \mathcal{I}). Each state $s \in \mathcal{S}$ is characterized by a pair $(w_s(\cdot), \mathcal{W}_s)$. Here, for each fixed state $s \in \mathcal{S}$,

$$w_s : \mathcal{I} \rightarrow \mathbb{R}_{++}$$

is a mapping assigning a positive stage payoff $w_s(i)$ to each individual $i \in \mathcal{I}$ ($w_s(i) > 0$ is useful as a normalization that makes zero payoff the worst outcome); \mathcal{W}_s is a (possibly empty) subset

of \mathcal{C} representing the set of *winning coalitions* in state s . Throughout the paper, we maintain the following assumption.

Assumption 1 (*Winning Coalitions*) For any state $s \in \mathcal{S}$, $\mathcal{W}_s \subset \mathcal{C}$ satisfies two properties:

- (a) If $X, Y \in \mathcal{C}$, $X \subset Y$, and $X \in \mathcal{W}_s$ then $Y \in \mathcal{W}_s$.
- (b) If $X, Y \in \mathcal{W}_s$, then $X \cap Y \neq \emptyset$.

Part (a) simply states that if some coalition X is winning for state s , then increasing the size of the coalition would not reverse this. This is a natural assumption for almost any decision rule. Part (b) rules out the possibility that two disjoint coalitions could be winning for the same state, thus imposing a form of (possibly weighted) majority or supermajority rule. Notice that $\mathcal{W}_s = \emptyset$ is not ruled out by this assumption. If $\mathcal{W}_s = \emptyset$, then state s is *exogenously stable*. For each state $s \in \mathcal{S}$, we also define the set of *blocking* coalitions as $\mathcal{B}_s = \{X \in \mathcal{C} \mid \mathcal{I} \setminus X \notin \mathcal{W}_s\}$. That is, a coalition is blocking if its complement is not winning. Clearly, $\mathcal{B}_s \subset \mathcal{W}_s$, meaning that winning coalitions are also blocking (but not necessarily vice versa). Blocking coalitions will be used in the noncooperative analysis in Section 4.

We define the following binary relations on the set of states \mathcal{S} . For $x, y \in \mathcal{S}$, we write

$$x \sim y \iff \forall i \in \mathcal{I} : w_x(i) = w_y(i). \quad (1)$$

In this case we call states x and y *payoff-equivalent*, or simply, *equivalent*. More important for our purposes is the binary relation \succeq_z . For any $z \in \mathcal{S}$, \succeq_z is defined by

$$y \succeq_z x \iff \{i \in \mathcal{I} : w_y(i) \geq w_x(i)\} \in \mathcal{W}_z. \quad (2)$$

Intuitively, $y \succeq_z x$ whenever a coalition of players that is winning (in z) weakly prefers y to x . If (2) holds, we say that y is *weakly preferred* to x in z (though naturally this does not mean that all individuals prefer y to x ; instead, a winning coalition of players does so). Relation \succ_z is defined by

$$y \succ_z x \iff \{i \in \mathcal{I} : w_y(i) > w_x(i)\} \in \mathcal{W}_z. \quad (3)$$

If (3) holds, we say that y is *strictly preferred* to x in z .

Relation \sim clearly defines an equivalence class, in that if $x \sim y$ and $y \sim z$, then $x \sim z$. In contrast, the binary relations \succeq_z and \succ_z may not even be transitive. Nevertheless, for any $x, z \in \mathcal{S}$, we have $x \not\succeq_z x$, and whenever \mathcal{W}_z is nonempty (that is, whenever z is not exogenously stable), we also have $x \succeq_z x$. Finally, for any $x, y, z \in \mathcal{S}$, $y \succ_z x$ implies $x \not\succeq_z y$, and similarly $y \succeq_z x$ implies $x \not\succeq_z y$; these implications follow from Assumption 1.

The following assumption introduces some basic properties of payoff functions and places some joint restrictions on payoff functions and winning coalitions.

Assumption 2 (Payoffs) *Payoff functions $\{w_s(\cdot)\}_{s \in \mathcal{S}}$ satisfy the following properties:*

(a) *For any nonempty collection of states $\mathcal{Q} \subset \mathcal{S}$, there exists state $z \in \mathcal{Q}$ such that for any $x \in \mathcal{Q}$, $x \not\succeq_z z$.*

(b) *For any nonempty collection of states $\mathcal{Q} \subset \mathcal{S}$ and for any $s \in \mathcal{S}$ such that $\forall x \in \mathcal{Q} : x \succ_s s$, there exists $z \in \mathcal{Q}$ such that for any $x \in \mathcal{Q}$, $x \not\succeq_s z$. Moreover, if $x \in \mathcal{Q}$ and $y \in \mathcal{S} : y \not\succeq_s s$, then $y \not\succeq_s x$.*

Assumption 2(a,b) play a major role in our analysis and ensure “acyclicity”. In the absence of these assumptions Condorcet-type cycles cannot be ruled out. Part (a) of Assumption 2 requires that within any collection of states there exists a state z such that the set of players that prefer another state is not sufficiently large (not winning in z). Another way of putting this assumption is as follows: There is no cycle of states such that for each of them there is a winning coalition of players who strictly prefer the previous state. Part (b) of Assumption 2 is a similar assumption, ruling out cycles among coalitions that are winning within state s . This part of the assumption states that in any collection of states that are winning within state s , there exists some z that is not less preferred (within s) than the other states in the collection. It also imposes the related requirement that if x is preferred to s and y is not, then y cannot be preferred to x . It can be seen easily that part (a) of Assumption 2 rules out cycles of the form $y \succ_z z, x \succ_y y, z \succ_x x$, while part (b) rules out cycles of the form $y \succ_s z, x \succ_s y, z \succ_s x$. Interestingly, neither of the two parts of Assumption 2 follows from the other. The roles of these assumptions are further illustrated by Examples 6 and 7 in Appendix B. These examples show the types of cycles that can arise when either 2(a) or 2(b) fails and how this would invalidate our main results.

We view Assumptions 1 and 2 as natural given the set of economic environments we are studying and assume that they hold throughout the paper. In addition, we will also sometimes impose an additional requirement, which is provided next.

Assumption 3 (Comparability) *For $x, y, z \in \mathcal{S}$ such that $x \succ_z z, y \succ_z z$, and $x \approx y$, either $y \succ_z x$ or $x \succ_z y$.*

This assumption states that if two states y and z are weakly preferred to x (in x), then y and z are \succ_x -comparable. It turns out that this condition is precisely the one necessary to guarantee uniqueness of equilibria. This assumption is not necessary for a range of our results, and for this reason, some of our main results are stated without imposing it. Example 8, also provided in Appendix B, shows how the absence of this assumption leads to multiplicity of equilibria. Nevertheless, even in that case we are able to characterize all these equilibria.

At each date, each individual maximizes her discounted expected utility:

$$U_t(i) = (1 - \beta) \sum_{\tau=t}^{\infty} \beta^{\tau-t} u_{\tau}(i), \quad (4)$$

where $\beta \in (0, 1)$ is a common discount factor and for now we can think of $u_t(i)$ as given by the payoff function $w_i(\cdot)$ introduced in Assumption 2. Throughout, we will consider situations in which β is greater than some threshold $\beta_0 \in (0, 1)$, which will be explicitly derived as a function of payoffs.

3 Axiomatic Characterization

The previous section described an economic and political environment; individuals derive utility from a state characterized by a particular set of rules, regulations, and ruling coalitions, and each state also specifies the distribution of political power and the political rules for determining future states (i.e., whether the society will remain in the same state or will undergo reform and transition to another state). The extensive-form game, specifying how voting decisions and transitions take place, will be presented in the next section. Before presenting this extensive-form game and its analysis, we first present an axiomatic characterization of “stable states”. This analysis has two purposes. First, it will show that the essential economic forces emphasized by our approach can be succinctly analyzed without specifying an explicit game form. Second, it will provide a characterization of the states that will then emerge as the dynamically stable states in the noncooperative game of the next section.

The key economic insight enabling an axiomatic characterization is the following: *with sufficiently forward-looking behavior, an individual should not wish to transition to a state that will ultimately lead to another state giving her lower utility.* This basic insight enables a tight characterization of (*axiomatically*) *stable states* (or simply stable states). The rationale for choosing this term is both because these states will be the stable points of a certain mapping ϕ , defined below. Moreover, Theorem 2 in the next section will show the equivalence between the notions of (*axiomatically*) stable states and the dynamically stable states of our noncooperative game.

Our axiomatic characterization will determine a mapping

$$\phi : \mathcal{S} \rightarrow \mathcal{S}$$

that will assign to each initial state $s_0 \in \mathcal{S}$ a dynamically stable state $s^{\infty} \in \mathcal{S}$. This axiomatic characterization will therefore bypass the analysis of the dynamics leading to this stable state, but simply determine the dynamically stable states given the initial state $s_0 \in \mathcal{S}$.

In the spirit of the discussion in the previous two paragraphs, we impose the following three axioms on this mapping.

Axiom 1 (*Desirability*) *If $x, y \in \mathcal{S}$ are such that $y = \phi(x)$, then either $y = x$ or $y \succ_x x$.*

Axiom 2 (*Rationality*) *If $x, y, z \in \mathcal{S}$ are such that $z \succ_x x$, $z = \phi(z)$, and $z \succ_x y$, then $y \neq \phi(x)$.*

Axiom 3 (*Stability*) *If $x, y \in \mathcal{S}$ are such that $y = \phi(x)$, then $y = \phi(y)$.*

All three axioms are natural in light of what we have discussed above. Axiom 1 essentially says that the population will not move to another state unless there is a winning coalition that supports this transition (for example, depending on the specification, $y \succ_x x$ might mean that starting with state x and majority rule, the majority of the population will vote for a reform towards y). Axiom 2 imposes the idea that if there exists some state z preferred to y by the group of decisive individuals starting in state x , then ϕ should not pick y ahead of z starting in x . Finally and most importantly, Axiom 3 encapsulates the stability notion discussed above—that an individual should not prefer a state that will ultimately lead to another, less preferred state. This notion is economically captured by the statement that if mapping ϕ will pick state y starting from state x , then it should also pick y starting from y (otherwise, y would lead to another state z , and as stated by Axiom 2, if this state z were indeed preferred to y , then ϕ would have picked z in the first instance).

It is important to emphasize that all three axioms should be thought of as properties of individual preferences. Then, because collective decision-making aggregates individual preferences, they indirectly apply to the mapping ϕ that summarizes these collective preferences (for example, one might think that ϕ aggregates individual preferences according to majority rule or weighted supermajority rule, and so on).

The next definition reiterates the meaning of dynamically stable states and its relationship to mapping ϕ .

Definition 1 (*Stable States*) *For any $\phi : \mathcal{S} \rightarrow \mathcal{S}$ that satisfies Axioms 1–3, a state $s \in \mathcal{S}$ is (axiomatically) stable if $\phi(s) = s$. The set of dynamically stable states is $\mathcal{D}_\phi = \{s \in \mathcal{S} : \phi(s) = s\}$.*

The next theorem is one of our main results. It establishes the existence of stable states and provides a recursive characterization of such states.

Theorem 1 (Axiomatic Characterization of Stable States) *Suppose Assumptions 1 and 2 hold. Then:*

1. *There exists a mapping ϕ satisfying Axioms 1–3.*
2. *Any ϕ that satisfies Axioms 1–3 can be recursively computed as follows. Let $\mu_1 \in \mathcal{S}$ be such that $\phi(\mu_1) = \mu_1$. Then, construct the sequence of states $\{\mu_1, \dots, \mu_{|\mathcal{S}|}\}$ with the property that if for any $l \in (j, |\mathcal{S}|]$, $\mu_l \not\prec_{\mu_j} \mu_j$. Let*

$$\mathcal{M}_k = \{s \in \{\mu_1, \dots, \mu_{k-1}\} : s \succ_{\mu_k} \mu_k \text{ and } \phi(s) = s\}.$$

Then, for each $k = 2, \dots, |\mathcal{S}|$,

$$\phi(\mu_k) = \begin{cases} \mu_k & \text{if } \mathcal{M}_k = \emptyset \\ s \in \mathcal{M}_k : \nexists z \in \mathcal{M}_k \text{ with } z \succ_{\mu_k} s & \text{if } \mathcal{M}_k \neq \emptyset \end{cases}.$$

(If there exist more than one $s \in \mathcal{M}_k$: $\nexists z \in \mathcal{M}_k$ with $z \succ_{\mu_k} s$, we pick any of these; this corresponds to multiple ϕ functions).

3. *For any two mappings ϕ_1 and ϕ_2 that satisfy Axioms 1–3 the stable states of these mappings coincide. In other words, $\mathcal{D}_{\phi_1} = \mathcal{D}_{\phi_2} = \mathcal{D}$.*
4. *If, in addition, Assumption 3 holds, then the mapping that satisfies Axioms 1–3 is “payoff-unique” in the sense that for any two mappings ϕ_1 and ϕ_2 that satisfy Axioms 1–3 and for any $s \in \mathcal{S}$, $\phi_1(s) \sim \phi_2(s)$.*

Proof. (Part 1) To prove existence, we first construct the sequence of states $\{\mu_1, \dots, \mu_{|\mathcal{S}|}\}$ such that

$$\text{if } 1 \leq j < l \leq |\mathcal{S}|, \text{ then } \mu_l \not\prec_{\mu_j} \mu_j, \quad (5)$$

The construction is by induction. Suppose we have defined μ_j for all $j \leq k-1$, where $k \leq |\mathcal{S}|$. Then applying Assumption 2(a) to the collection of states $\mathcal{S} \setminus \{\mu_1, \dots, \mu_{k-1}\}$, we conclude that there exists μ_k satisfying (5). By construction, μ is a bijection that satisfies (5).

The second step is to construct a candidate mapping $\phi : \mathcal{S} \rightarrow \mathcal{S}$. This is again by induction. For $k = 1$, let $\phi(\mu_k) = \mu_k$. Suppose we have defined μ_j for all $j \leq k-1$ where $k \leq |\mathcal{S}|$. Define the collection of states

$$\mathcal{M}_k = \{s \in \{\mu_1, \dots, \mu_{k-1}\} : s \succ_{\mu_k} \mu_k \text{ and } \phi(s) = s\}. \quad (6)$$

\mathcal{M}_k is the subset of states where ϕ has already been defined and satisfies $\phi(s) = s$ and which are preferred to μ_k within μ_k . If \mathcal{M}_k is empty, then we define $\phi(\mu_k) = \mu_k$. If \mathcal{M}_k is non-empty, then take $\phi(\mu_k)$ such that

$$s \not\prec_{\mu_k} \phi(\mu_k) \text{ for any } s \in \mathcal{M}_k \quad (7)$$

(such state $\phi(\mu_k)$ exists because we can apply Assumption 2(b) to \mathcal{M}_k). Proceeding likewise for all $2 \leq k \leq |\mathcal{S}|$ we construct mapping ϕ .

To complete the proof, we need to verify that mapping ϕ satisfies Axioms 1–3. This is straightforward for Axioms 1 and 3. In particular, by construction, either $\phi(\mu_k) = \mu_k$ (in that case these axioms trivially hold), or $\phi(\mu_k)$ is an element of \mathcal{M}_k ; in that case $\phi(\mu_k) \succ_{\mu_k} \mu_k$ and $\phi(\phi(\mu_k)) = \phi(\mu_k)$ by (6). To check Axiom 2, suppose that for some state μ_k there exists z such that $z \succ_{\mu_k} \mu_k$, $z = \phi(z)$, and $z \succ_{\mu_k} \phi(\mu_k)$. Then $z \succ_{\mu_k} \mu_k$, combined with condition (5), implies that $z \in \{\mu_1, \dots, \mu_{k-1}\}$. But the last condition, $z \succ_{\mu_k} \phi(\mu_k)$, now contradicts (7). This means that such z does not exist, and therefore Axiom 2 is satisfied.

(Part 2) Suppose that ϕ_1 and ϕ_2 are two mappings satisfying Axioms 1–3. Consider the sequence of states $\{\mu_k\}_{k=1}^{|\mathcal{S}|}$. If the sets of stable states of ϕ_1 and ϕ_2 do not coincide, there exists some k such that $\phi_1(\mu_j) = \mu_j \iff \phi_2(\mu_j) = \mu_j$ for $j < k$, but either $\phi_1(\mu_k) = \mu_k$ and $\phi_2(\mu_k) \neq \mu_k$ or $\phi_1(\mu_k) \neq \mu_k$ and $\phi_2(\mu_k) = \mu_k$. Without loss of generality assume the former is the case (the argument for the latter case is identical). By Axiom 1, $\phi_2(\mu_k) \succ_{\mu_k} \mu_k$, so by (5) $\phi_2(\mu_k) = \mu_l$ for some $l < k$. Therefore, we have $\mu_l \succ_{\mu_k} \mu_k$ and because $l < k$ and $\phi_1(\mu_l) = \phi_2(\mu_l) = \mu_l$, $\phi_1(\phi_1(\mu_l)) = \phi_1(\phi_2(\mu_l)) = \phi_1(\mu_l)$. But then Axiom 2 implies that $\phi_1(\mu_k)$ cannot equal y which satisfies $\mu_l \succ_{\mu_k} y$, in particular, $\phi_1(\mu_k) \neq \mu_k$, yielding a contradiction to the hypothesis that $\phi_1(\mu_k) = \mu_k$ and $\phi_2(\mu_k) \neq \mu_k$. Consequently, $\phi_1(s) = s$ if and only if $\phi_2(s) = s$.

(Part 3) Suppose Assumption 3 holds. Suppose, to obtain a contradiction, that ϕ_1 and ϕ_2 are two distinct mappings that satisfy Axioms 1–3. Then there exists some state s such that $\phi_1(s) \approx \phi_2(s)$. Part 2 of this Theorem implies that $\phi_1(s) = s$ if and only if $\phi_2(s) = s$; since $\phi_1(s) \approx \phi_2(s)$, we obtain that $\phi_1(\mu_{ks}) \neq s \neq \phi_2(s)$. Now Axiom 1 implies $\phi_1(s) \succ_s s$, $\phi_2(s) \succ_s s$, and Assumption 3 implies that either $\phi_1(s) \succ_s \phi_2(s)$ or $\phi_2(s) \succ_s \phi_1(s)$; without loss of generality assume the former. Then for $y = \phi_2(s)$ there exists $\phi_1(s)$ such that $\phi_1(s) \succ_s y$, $\phi_1(s) \succ_s s$, and $\phi_2(\phi_1(s)) = \phi_1(s)$ (the latter because $\phi_1(s)$ is a ϕ_1 -stable state by Axiom 3, and by Part 2 of this Theorem, it is also a ϕ_2 -stable state). But then we can apply Axiom 2 to mapping ϕ_2 and derive the conclusion that $\phi_2(s)$ cannot equal y . This contradiction completes the proof. ■

Theorem 1 shows that a mapping that satisfies Axioms 1–3 necessarily exists and provides a sufficient condition for its uniqueness. Even when the uniqueness condition, Assumption 3, does not hold, we know that axiomatically stable states coincide for any two mappings ϕ_1 and ϕ_2 that satisfy Axioms 1–3.

Theorem 1 also provides a simple recursive characterization of the mapping ϕ . Intuitively,

Assumption 2(a) ensures that there exists some state $\mu_1 \in \mathcal{S}$, such that there does not exist another state $s \in \mathcal{S}$ with $s \succ_{\mu_1} \mu_1$. Taking μ_1 as base, we recursively construct the set of states $\mathcal{M}_k \subset \mathcal{C}$, $k = 1, \dots, |\mathcal{S}|$, that includes (axiomatically) stable states that are preferred to state μ_k (that is, $\phi(s) = s$ and $s \succ_{\mu_k} \mu_k$). When the set \mathcal{M}_k is empty, then no stable state is part of a winning coalition starting in state μ_k , and therefore we must have $\phi(\mu_k) = \mu_k$. When this set is nonempty, then we can pick a stable state that will arise starting from state μ_k . In addition to its recursive (and thus easy-to-construct) nature, this characterization is useful because it highlights the fundamental property of stable states emphasized in the Introduction: *a state μ_k is made stable precisely by the absence winning coalitions in μ_k favoring a transition to another stable state.* We will see that this insight plays an important role in the applications in Section 6.

We have motivated the analysis leading up to Theorem 1 with the argument that, when agents are sufficiently forward-looking, only axiomatically stable states should be observed (at least in the “long run”). The analysis of the extensive-form game in the next section will substantiate this interpretation further.

4 Noncooperative Foundations of Dynamically Stable States

We now describe an extensive-form game meant to capture the basic economic interactions emphasized so far in the simplest possible way. The main result of this section will establish the equivalence between the Markov Perfect Equilibria (MPE) of this extensive-form game and the axiomatic characterization of Theorem 1.

The essential features of the extensive-form game are: (i) a protocol for a sequence of agenda-setters and proposals at each date; and (ii) a protocol for voting over proposals. The protocol for voting is taken to be sequential voting and is described below. We represent the protocol for agenda-setting using a mapping π_s . Let K_s be a natural number. Then,

$$\pi_s : \{1, \dots, K_s\} \rightarrow \mathcal{I} \cup (\mathcal{S})$$

for each state $s \in \mathcal{S}$. This mapping specifies a finite sequence of elements from $\mathcal{I} \cup \mathcal{S}$, where K_s ($\leq |\mathcal{I}| + |\mathcal{S}|$) is the length of sequence for state s and determines the sequence of agenda-setters and proposals. In particular, if $\pi_s(\tau) \in \mathcal{I}$, then it denotes an agenda-setter who will make a proposal from the set of states \mathcal{S} . Alternatively, if $\pi_s(\tau) \in \mathcal{S}$, then it directly corresponds to an exogenously-specified proposal over which individuals vote. Therefore, the extensive-form game is general enough to include both proposals for a change to a new state initiated by agenda-setters (a subset of the players that may depend on the state) or those that are exogenously

placed on the table (as is the case in standard voting models where alternatives are voted over in pairwise contests). We impose the following mild requirements on $\pi_s(\cdot)$:

Assumption 4 (*Agenda-Setting and Proposals*) For every state $s \in \mathcal{S}$, one (or both) of the following two conditions is satisfied:

- (a) For any state $q \in \mathcal{S} \setminus \{s\}$, there is an element $k : 1 \leq k \leq K_s$ of sequence π_s such that $\pi_s(k) = q$.
- (b) For any player $i \in \mathcal{I}$ there is an element $k : 1 \leq k \leq K_s$ of sequence π_s such that $\pi_s(k) = i$.

This assumption implies that either sequence π_s contains all possible states (other than “status quo” s) as proposals, or it allows all possible agenda-setters to eventually make a proposal. It ensures that either all alternatives will be considered, or all players will have a chance to propose.

At $t = 0$, state $s_0 \in \mathcal{S}$ is taken as given (as noted above, it might be determined as part of the description of the environment or determined by Nature according to some probability distribution). In each period starting from $t = 1$, there is a finite sequence of votings over proposals to change the current state s . The sequence is determined by π as described in Assumption 4. Within any period t , if a proposal s' receives sufficient support, then a transition takes place from s_{t-1} to s' . If a proposal is turned down, then the next proposal is considered. If all proposals are turned down, then the period ends and the next period begins with the same state s .

More precisely, the timing within each period $t \geq 1$ is as follows:

1. Period t begins with state s_{t-1} inherited from the previous period.
2. For $k = 1, \dots, K_{s_{t-1}}$, the k th proposal $P_{k,t}$ is determined as follows. If $\pi_{s_{t-1}}(k) \in \mathcal{S}$, then $P_{k,t} = \pi_{s_{t-1}}(k)$. If $\pi_{s_{t-1}}(k) \in \mathcal{I}$, then player $\pi_{s_{t-1}}(k)$ chooses $P_{k,t} \in \mathcal{S}$.
3. If $P_{k,t} \neq s_{t-1}$, then there is a sequential voting between two alternatives, $P_{k,t}$ and s_{t-1} (we will show that the sequence in which voting takes place has no effect on the equilibrium and we do not specify it here). Each player votes *yes* (for $P_{k,t}$) or *no* (for s_{t-1}). Let $Y_{k,t}$ denote the set of players who voted *yes*. If $Y_{k,t} \in \mathcal{W}_{t-1}$, then alternative $P_{k,t}$ is accepted, otherwise (that is, if $Y_{k,t} \notin \mathcal{W}_s$), it is rejected. If $P_{k,t} = s_{t-1}$, there is no voting and we adopt the convention that in this case $P_{k,t}$ is rejected.
4. If $P_{k,t}$ is accepted in the voting, then the next state $s_t = P_{k,t}$, and the period ends. If $P_{k,t}$ is rejected, or if there is no voting because $P_{k,t} = s_{t-1}$, then the game moves to step 2

with k increased by 1 as long as $k < K_{s_{t-1}}$. If $k = K_{s_{t-1}}$, the next state is $s_t = s_{t-1}$, and this period ends.

5. In the end of the period, each player receives instantaneous utility $u_t(i)$.

Payoffs in this dynamic game are given by (4), with

$$u_t(i) = \begin{cases} w_s(i) & \text{if } s_t = s_{t-1} = s \\ 0 & \text{if } s_t \neq s_{t-1} \end{cases}$$

for each $i \in \mathcal{I}$. In other words, in the period in which a transition occurs (that is, if the current state is not the same as the state in the previous period), each individual receives zero payoff. In all other periods, each individual receives the payoff as specified in Assumption 2. The period of zero payoff can be interpreted as representing a “transaction cost” associated with the change in the state and is introduced to guarantee the existence of pure-strategy MPE. Since the game is infinitely-repeated and we will take β to be large, this one period of “transaction cost” typically has little effect on discounted payoffs. In particular, once (and if) a dynamically stable state is reached, individuals will receive $w_s(i)$ at each date thereafter.⁴

We next define a Markov Perfect Equilibrium (MPE). Loosely speaking, MPE is the subset of Subgame Perfect Equilibria where strategies are only functions of “payoff-relevant states,” where “payoff-relevant states” are different from the states $s \in S$ described above, since the order in which proposals have been made within a given period are also payoff relevant for the continuation game. To define MPE more formally, consider a general n -person infinite-stage game, where each individual can take an action at every stage. Let the action profile of each individual be $a^i = (a_1^i, a_2^i, \dots)$ for $i = 1, \dots, n$, with $a_t^i \in A_t^i$ and $a^i \in A^i = \prod_{t=1}^{\infty} A_t^i$. Let $h_t = (a_1, \dots, a_t)$ be the history of play up to stage t (not including stage t), where $a_s = (a_s^1, \dots, a_s^n)$, so h_0 is the history at the beginning of the game, and let H_t be the set of histories h_t for $t : 0 \leq t \leq T-1$. We denote the set of all potential histories up to date t by $H^t = \bigcup_{s=0}^t H_s$. Let t -continuation action profiles be $a^{i,t} = (a_t^i, a_{t+1}^i, \dots)$ for $i = 1, \dots, n$, with the set of continuation action profiles for player i denoted by $A^{i,t}$. Symmetrically, define t -truncated action profiles as $a^{i,-t} = (a_1^i, a_2^i, \dots, a_{t-1}^i)$ for $i = 1, \dots, n$, with the set of t -truncated action profiles for player i denoted by $A^{i,-t}$. We also use the standard notation a^i and a^{-i} to denote the action profiles for player i and the action profiles of all other players (similarly, A^i and A^{-i}). The payoff functions for the players depend only on actions, i.e., player i 's payoff is given by $u^i(a^1, \dots, a^n)$. A pure

⁴It should also be noted that various different alternative game forms lead to same results, and we present the current one because it appears to correspond both the simplest and the most natural protocols for agenda setting and voting. In particular, it allows votes to take place over all possible proposals (or all possible agenda-setters have a move), which is a desirable feature, since otherwise some transitions would be ruled out by the game form.

strategy for player i is

$$\sigma^i : H^\infty \rightarrow A^i.$$

A t -continuation strategy for player i (corresponding to strategy σ^i) specifies plays only after time t (including time t), i.e.,

$$\sigma^{i,t} : H^\infty \setminus H^{t-2} \rightarrow A^{i,t},$$

where $H^\infty \setminus H^{t-2}$ is the set of histories starting at time t . We then have:

Definition 2 (Markovian Strategies) A continuation strategy $\sigma^{i,t}$ is Markovian if

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

for all $\tau \geq t$, whenever $h^{t-1}, \tilde{h}^{\tau-1} \in H^\infty$ are such that for any $a^{i,t}, \tilde{a}^{i,\tau} \in A^{i,t}$ and any $a^{-i,t} \in A^{-i,t}$,

$$u^i(a^{i,t}, a^{-i,t} | h^{t-1}) \geq u^i(\tilde{a}^{i,\tau}, a^{-i,t} | h^{\tau-1})$$

implies

$$u^i(a^{i,t}, a^{-i,t} | \tilde{h}^{t-1}) \geq u^i(\tilde{a}^{i,\tau}, a^{-i,t} | \tilde{h}^{\tau-1}).$$

Definition 3 (MPE) A pure strategy profile $\hat{\sigma} = (\hat{\sigma}^1, \dots, \hat{\sigma}^n)$ is Markov Perfect Equilibrium (MPE) (in pure strategies) if each strategy $\hat{\sigma}^i$ is Markovian and

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

In what follows, we will use the terms MPE and equilibrium interchangeably. We next define dynamically stable states.

Definition 4 (Dynamically Stable States) State $s^\infty \in \mathcal{S}$ is a dynamically stable state if there exist a set of sequences $\{\pi_s(\cdot)\}_{s \in \mathcal{S}}$, a MPE strategy profile σ (for a game starting with initial state s_0) and $T < \infty$, such that along the equilibrium path we have $s_t = s^\infty$ for all $t \geq T$.

Put differently, s^∞ is a dynamically stable state if it is reached by time T and is repeated thereafter. Our objective is to determine whether dynamically stable states exist in the extensive-form game described above and to characterize these dynamically stable states as a function of the initial state $s_0 \in \mathcal{S}$. We will then also establish the equivalence between dynamically stable states and the axiomatically stable states characterized in the previous section.

Our main result is given in the following theorem and establishes a close correspondence between the MPEs of the game described here and the outcomes picked by mapping ϕ described in Theorem 1. For this result, we also introduce a slightly stronger version of Assumption 2(b).

Assumption 2(b)* For any nonempty collection of states $\mathcal{Q} \subset \mathcal{S}$ such that not all states in \mathcal{Q} are payoff-equivalent and for any $s \in \mathcal{S}$ such that $\forall x \in \mathcal{Q} : x \succ_s s$, there exists $z \in \mathcal{Q}$ such that for any $x \in \mathcal{Q} \setminus \{z\}$, $x \not\prec_s z$. Moreover, if $x \in \mathcal{Q}$ and $y \in \mathcal{S} : y \not\prec_s s$, then $y \not\prec_s x$.

Notice that this assumption has substituted the requirement that $x \not\prec_s z$ in Assumption 2(b) by $x \not\prec_s z$. Therefore, in addition to cycles of the form $x \succ_s y, y \succ_s z, z \succ_s x$ ruled out by Assumption 2(b), this assumption rules out cycles of the form $x \succeq_s y, y \succeq_s z, z \succeq_s x$, unless the states x, y , and z are payoff-equivalent. We now have:

Theorem 2 (Noncooperative Foundations of Dynamically Stable States) Suppose Assumptions 1 and 2(a),(b). Then there exists $\beta_0 \in [0, 1)$ such if the discount factor $\beta \geq \beta_0$, then:

1. For any mapping ϕ satisfying Axioms 1–3 there is a set of sequences $\{\pi_s(\cdot)\}_{s \in \mathcal{S}}$ and a MPE σ of the game such that $s_t = \phi(s_0)$ for any $t \geq 1$; that is, the game reaches $\phi(s_0)$ after one period and stays in this state thereafter. Therefore, $s = \phi(s_0)$ is a dynamically stable state.

Moreover, suppose that Assumption 2(b)* holds. Then:

2. For any set of sequences $\{\pi_s(\cdot)\}_{s \in \mathcal{S}}$, any MPE in pure strategies σ has the property that for any initial state $s_0 \in \mathcal{S}$, it reaches some state, s^∞ , in a finite number of periods (with at most one transition), so that for $t \geq 1$, $s_t = s^\infty$. Moreover, there exists mapping $\phi : \mathcal{S} \rightarrow \mathcal{S}$ that satisfies Axioms 1–3 such that $s^\infty = \phi(s_0)$. Therefore, all dynamically stable states are axiomatically stable.
3. If, in addition, Assumption 3 holds, then the MPE is essentially unique in the sense that for any set of sequences $\{\pi_s(\cdot)\}_{s \in \mathcal{S}}$, any MPE strategy profile in pure strategies σ induces $s_t \sim \phi(s_0)$ for all $t \geq 1$, where ϕ satisfies Axioms 1–3.

Proof. See Appendix A. ■

Parts 1 and 2 of 2 may be summarized as follows: the set of dynamically stable states and the set of stable states \mathcal{D} defined by axiomatic characterization in Theorem 1 coincide; any mapping ϕ satisfying Axioms 1–3 is implementable by a Markov Perfect equilibrium, and any MPE implements a mapping that satisfies Axioms 1–3. This theorem therefore establishes the equivalence of axiomatic and dynamic characterizations and enables us to use all of the properties of the axiomatically stable states (in particular, their recursive characterization) in the context of the noncooperative game presented in this section.

The equivalence on the results of Theorems 1 and 2 is intuitive. Had the players been short-sighted (impatient), they would care mostly about the payoffs in the next state or the next few states that would arise along the equilibrium path (as in the concept of myopic stability introduced next). However, when players are sufficiently patient, in particular, when $\beta \geq \beta_0$, they care more about payoffs in the ultimate state that the equilibrium path will lead to. Consequently, winning coalitions will be not willing to move to a state that is not (axiomatically) stable according to Theorem 1, and this leads to the equivalence between the concepts of axiomatically and dynamically stable states.

To highlight some of the implications of our analysis so far and emphasize the difference between dynamically stable states and states that may arise when individuals are shortsighted, we next introduce a number of corollaries of Theorems 1 and 2. These corollaries will be particularly useful in highlighting the two general messages of our approach discussed in the Introduction. We start with a simple definition.

Definition 5 (*Myopic Stability*) *A state $s^m \in \mathcal{S}$ is myopically stable if there does not exist $s' \in \mathcal{S}$ with $s' \succ_{s^m} s^m$.*

Essentially, myopic stability would apply if individuals made choices only considering the implications in the next period. Clearly, a myopically stable state is (axiomatically and dynamically) stable, but the converse is not true. This is stated in the next corollary, which emphasizes the important feature that a state is made stable not by the absence of a powerful group preferring change, but by the absence of an alternative stable state that is preferred by a powerful group. This corollary is an immediate implication of Theorems 1 and 2, and thus its proof is omitted.

Corollary 1 *1. State $s^\infty \in \mathcal{S}$ is a (dynamically and axiomatically) stable state only if for any $s' \in \mathcal{S}$ with $s' \succ_{s^\infty} s^\infty$, and any ϕ satisfying Axioms 1–3, $s' \neq \phi(s')$.*

2. A myopically stable state s^m is a stable state.

3. A stable state s^∞ is not necessarily myopically stable.

The final part of the corollary implies that s^∞ may be stable in general despite the fact that it may not be myopically stable, that is, there may exist states s' such that $s' \succ_{s^\infty} s^\infty$. State s^∞ is nonetheless stable, because these alternative states themselves are not stable (recall, for instance, the simple illustration in Example 1). Intuitively, alternative states such as s' described in this paragraph do not arise in equilibrium because, by the fact that $s' \neq \phi(s')$, they involve a

change in the distribution of political power and thus would lead to some other state s'' , which is not preferred by a winning coalition in s^∞ (if we had $s'' \succ_{s^\infty} s^\infty$, then s^∞ would not have been a stable state).

For the next corollary, we first introduce an additional definition.

Definition 6 (Inefficiency) *State $s \in S$ is (strictly) Pareto inefficient if there exists a state $s' \in S$ such that $w_{s'}(i) > w_s(i)$ for all $i \in \mathcal{I}$.*

State $s \in S$ is (strictly) winning coalition inefficient if there exists a winning coalition $\mathcal{W}_s \subset \mathcal{I}$ in s and $s' \in S$ such that $w_{s'}(i) > w_s(i)$ for all $i \in \mathcal{W}_s$.

Clearly, if a state s is Pareto inefficient, it is winning coalition inefficient, but not vice versa.

Corollary 2 *1. A stable state $s^\infty \in S$ can be (strictly) winning coalition inefficient and Pareto inefficient.*

2. Whenever s^∞ is not myopically stable, it is winning coalition inefficient.

Proof. The first part again follows from Example 1 in the Introduction. The second part follows from the fact that if s^∞ is not myopically stable, then there must exist $s' \in S$ such that $s' \succ_{s^\infty} s^\infty$. ■

5 Limited State Transitions

We have so far assumed that any transition (from any state into any other state) is possible. In many interesting applications, there will be certain transitions that are not possible. For example, in Example 1 discussed in the Introduction, it may be that a transition to democracy is only possible from constitutional monarchy (and not directly from absolutist monarchy). Another more substantial example highlighting the importance of limited transitions is the model in Acemoglu, Egorov, and Sonin (2008) mentioned in Example 4 in the Introduction and discussed in greater detail in subsection 6.5. In that model, only current members of the ruling coalition can be part of future ruling coalitions and thus transitions to states that include individuals previously eliminated are ruled out. In this section we reformulate the Assumptions and the Axioms of Section 2 to incorporate the feature that only certain state transitions are allowed and generalize the results in Theorems 1 and 2.

The key to the analysis in this section is the binary relation \rightsquigarrow on the set of states \mathcal{S} . For any $x, y \in \mathcal{S}$, we write $x \rightsquigarrow y$ to denote that a transition from x to y is possible and $x \rightsquigarrow \mathcal{Q}$ for some $\mathcal{Q} \subset \mathcal{S}$ to denote that the transition to any state z in \mathcal{Q} is possible (provided that these

positions are supported by a winning coalition in x). Our analysis so far thus corresponds to the special case where $x \rightsquigarrow \mathcal{S}$ for any $x \in \mathcal{S}$. We adopt the following natural assumption on the transition relation.

Assumption 5 (*Feasible Transitions*) Relation \rightsquigarrow satisfies the following properties:

- (a) (*reflexivity*) $\forall x \in \mathcal{S} : x \rightsquigarrow x$;
- (a) (*transitivity*) $\forall x, y, z \in \mathcal{S} : x \rightsquigarrow y$ and $y \rightsquigarrow z$ imply $x \rightsquigarrow z$.

Part (b) Assumption 5 requires that if some indirect transition from x to z is feasible, so is a direct transition between the states. Without requiring transitivity, there would be additional technical details to take care of, because, for instance, if transition from x to z is possible through y only, then it is only possible if both a winning coalition in x prefers z to x and a winning coalition in y prefers z to y .⁵ Nevertheless, this assumption can be dispensed with, and we could assume instead that whenever $x \rightsquigarrow y$ and $y \rightsquigarrow z$ but $x \not\rightsquigarrow z$, then $\mathcal{W}_x = \mathcal{W}_y$ (or a weaker version of this assumption).⁶

We next consider slightly weaker versions of Assumption 2 and Assumption 3, incorporating the fact that only certain transitions are feasible (since when some transitions are not feasible, it becomes easier to rule out cycles), and we also introduced the equivalent of Assumption 2(b)* here.

Assumption 2' (*Payoffs with Limited Transitions*) Payoff functions $\{w_s(\cdot)\}_{s \in \mathcal{S}}$ satisfy the following properties:

- (a) For any nonempty $\mathcal{Q} \subset \mathcal{S}$ such that $x \rightsquigarrow y$ for any $x, y \in \mathcal{Q}$, there exists state $z \in \mathcal{Q}$ such that for any $x \in \mathcal{Q}$, $x \not\rightsquigarrow_z z$;
- (b) For any nonempty collection of states $\mathcal{Q} \subset \mathcal{S}$ and for any $s \in \mathcal{S}$ such that $s \rightsquigarrow x$ and $x \succ_s s$ for any $x \in \mathcal{Q}$ there exists $z \in \mathcal{Q}$ such that for any $x \in \mathcal{Q}$, $x \not\rightsquigarrow_s z$. Moreover, if $x \in \mathcal{Q}$ and $y \in \mathcal{S} : y \not\rightsquigarrow_s s$, and $s \rightsquigarrow x$, $s \rightsquigarrow y$, then $y \not\rightsquigarrow_s x$;
- (b)* For any nonempty collection of non-payoff-equivalent states $\mathcal{Q} \subset \mathcal{S}$ and for any $s \in \mathcal{S}$ such that $s \rightsquigarrow x$ and $x \succ_s s$ for any $x \in \mathcal{Q}$, there exists $z \in \mathcal{Q}$ such that for any $x \in \mathcal{Q}$, $x \not\rightsquigarrow_s z$. Moreover, $x \in \mathcal{Q}$ and $y \in \mathcal{S} : y \not\rightsquigarrow_s s$, and $s \rightsquigarrow x$, $s \rightsquigarrow y$, then $y \not\rightsquigarrow_s x$.

⁵See Chwe (1994) for another model where different transitions require different winning coalitions.

⁶One set of economically interesting cases in which Assumption 5 would fail to hold includes economic games in which there is a capital-stock like variable, such as capital, that is determined as a result of the actions in the current state (for example, capital accumulation, which might depend on the current enforcement of property rights). Since our game does not involve such dynamic linkages, Assumption 5 is natural here. In particular, there is no reason for a sufficiently powerful coalition not to be able to implement a change that is feasible in the continuation game. An interesting model of a gradual dynamic enfranchisement where capital accumulation changes agents' preferences over time is provided in Jack and Lagunoff (2006).

Assumption 3' (Comparability with Limited Transitions) For $x, y, z \in \mathcal{S}$ such that $z \rightsquigarrow x$, $z \rightsquigarrow y$, $x \succ_z z$, $y \succ_z z$, and $x \approx y$, either $y \succ_z x$ or $x \succ_z y$.

Finally, let us reformulate Axioms 1–3 for this slightly modified set up (note that Axiom 3 is unchanged, though we state it again for completeness).

Axiom 1' (Desirability) If $x, y \in \mathcal{S}$ are such that $y = \phi(x)$, then either $y = x$ or $x \rightsquigarrow y$ and $y \succ_x x$.

Axiom 2' (Rationality) If $x, y, z \in \mathcal{S}$ are such that $x \rightsquigarrow z$, $z \succ_x x$, $z = \phi(z)$, and $z \succ_x y$, then $y \neq \phi(x)$.

Axiom 3' (Stability) If $x, y \in \mathcal{S}$ are such that $y = \phi(x)$, then $y = \phi(y)$.

With this new set of Axioms, a slightly modified version of Theorem 1 holds:

Theorem 3 (Dynamically Stable States with Limited Transitions) Suppose that binary relation \rightsquigarrow satisfies Assumption 5, and that Assumptions 1 and 2' hold. Then:

1. There exists mapping ϕ satisfying Axioms 1'–3'.
2. Any mapping ϕ that satisfies Axioms 1'–3' this characterize recursively as follows. Construct sequence $\{\mu_1, \dots, \mu_{|\mathcal{S}|}\}$ with the property that if for any $l \in (j, |\mathcal{S}|]$, either $\mu_j \not\succeq \mu_l$ or $\mu_l \not\succeq \mu_j$. Let $\phi(\mu_1) = \mu_1$. For any $k : 2 \leq k \leq |\mathcal{S}|$, define

$$\mathcal{M}_k = \{s \in \{\mu_1, \dots, \mu_{k-1}\} : \mu_k \rightsquigarrow s, s \succ_{\mu_k} \mu_k, \text{ and } \phi(s) = s\}$$

and let

$$\phi(\mu_k) = \begin{cases} \mu_k & \text{if } \mathcal{M}_k = \emptyset \\ s \in \mathcal{M}_k : \nexists z \in \mathcal{M}_k \text{ with } \mu_k \rightsquigarrow z \text{ and } z \succ_{\mu_k} s & \text{if } \mathcal{M}_k \neq \emptyset \end{cases}.$$

3. For any two mappings ϕ_1 and ϕ_2 that satisfy Axioms 1'–3' the stable states of these mappings coincide.
4. If, in addition, Assumption 3' holds, then the mapping that satisfies Axioms 1'–3' is “payoff-unique” in the sense that for any two mappings ϕ_1 and ϕ_2 that satisfy Axioms 1'–3' and for any $s \in \mathcal{S}$, $\phi_1(s) \sim \phi_2(s)$.

Proof. The proof is an extension of that of Theorem 1. The idea of the proof is to construct a mapping (sequence) $\mu : \{1, \dots, |\mathcal{S}|\} \leftrightarrow \mathcal{S}$ such that for any $1 \leq k < |\mathcal{S}|$ we have that

$$\text{if } 1 \leq j < l \leq |\mathcal{S}|, \text{ then } \mu_j \not\prec \mu_l \text{ or } \mu_l \not\prec_{\mu_j} \mu_j. \quad (8)$$

To construct mapping μ , for each $x \in \mathcal{S}$ we consider its equivalence class \mathcal{E}_x defined by

$$\mathcal{E}_x = \{y \in \mathcal{S} : x \rightsquigarrow y \text{ and } y \rightsquigarrow x\}.$$

Assumption 5 guarantees that $\{\mathcal{E}_x \mid x \in \mathcal{S}\}$ indeed defines an equivalence relation with different classes either coinciding or not intersecting. The binary relation \rightsquigarrow on elements of \mathcal{S} induces relation \rightsquigarrow in equivalence classes by letting $\mathcal{E}_x \rightsquigarrow \mathcal{E}_y$ if and only if $x \rightsquigarrow y$; note that this relation is well-defined in the sense that it does not depend on the elements x and y picked from \mathcal{E}_x and \mathcal{E}_y , respectively. Furthermore, this relation is acyclical in the sense that there do not exist distinct classes $\mathcal{E}_1, \dots, \mathcal{E}_l$ such that $\mathcal{E}^j \rightsquigarrow \mathcal{E}^{j+1}$ for $1 \leq j < l$ and $\mathcal{E}^l \rightsquigarrow \mathcal{E}^1$. Consequently, we can form a sequence of all equivalence classes $\mathcal{E}^1, \dots, \mathcal{E}^m$ (where m is the number of classes) such that $\mathcal{E}^j \not\rightsquigarrow \mathcal{E}^k$ for any $1 \leq j < k \leq m$. Now, within each class \mathcal{E}^k , we enumerate its elements as $\mu_1^k, \dots, \mu_{|\mathcal{E}^k|}^k$ so that $\mu_l^k \not\prec_{\mu_j^k} \mu_j^k$ for $1 \leq j < l \leq |\mathcal{E}^k|$ (this is feasible due to Assumption 2'(a)). Next, construct the sequence μ as follows: we give members of class \mathcal{E}_1 numbers 1 to $|\mathcal{E}_1|$ in the order they are listed in the sequence $\mu^1 \equiv (\mu_1^1, \dots, \mu_{|\mathcal{E}_1|}^1)$, then we take members of class \mathcal{E}_2 as they are listed in the sequence μ^2 , and so on. It is easy to show that the sequence μ constructed in this way satisfies (8). The rest of the proof closely follows the one of Theorem 1 and is omitted. ■

Similarly, an equivalent of Theorem 2 again applies.

Theorem 4 (Noncooperative Foundations of Dynamically Stable States with Limited Transitions) *Suppose that binary relation \rightsquigarrow satisfies Assumption 5, that Assumptions 1 and 2'(a) hold. Then there exists $\beta_0 \in [0, 1)$ such if the discount factor $\beta \geq \beta_0$, then:*

1. *For any mapping ϕ satisfying Axioms 1'–3' there is a set of sequences $\{\pi_s(\cdot)\}_{s \in \mathcal{S}}$ and a MPE σ of the game such that $s_t = \phi(s_0)$ for any $t \geq 1$. In other words, the game reaches $\phi(s_0)$ in a finite number of steps (after one period) and stays in this state thereafter.*

Moreover, suppose that Assumption 2'(b) holds. Then:*

2. *For any set of sequences $\{\pi_s(\cdot)\}_{s \in \mathcal{S}}$, any MPE in pure strategies σ has the property that for any initial state $s_0 \in \mathcal{S}$ it reaches a certain state, s^∞ , in a finite number of periods (with at most one transition): for $t \geq 1$, $s_t = s^\infty$. Moreover, there exists mapping $\phi : \mathcal{S} \rightarrow \mathcal{S}$ that satisfies Axioms 1'–3' such that $s^\infty = \phi(s_0)$.*

3. If, in addition, Assumption 3' holds, then the MPE is essentially unique in the sense that for any set of sequences $\{\pi_s(\cdot)\}_{s \in \mathcal{S}}$, any MPE strategy profile in pure strategies σ induces $s_t \sim \phi(s_0)$ for all $t \geq 1$, where ϕ satisfies Axioms 1–3.

Proof. The proof is essentially identical to that of Theorem 2 and is omitted. ■

These theorems therefore show that the essential results of Theorems 1 and 2 generalize to an environment with limited transitions. The intuition for these results and the recursive characterization of dynamically stable states are essentially identical to those in Theorems 1 and 2.

6 Applications

We now revisit the examples discussed in the Introduction, as well as a number of new examples, and show how the theory developed above can be applied in these cases to derive predictions about dynamically stable states. In some of the applications, we will allow for $w_s(i) = 0$ for some i and s . This is to simplify notation, and setting $w_s(i) = \varepsilon$ for $\varepsilon > 0$ and small would not change any of the results or interpretations.

6.1 Inefficient Inertia and Lack of Reform

We now provide a more detailed example capturing the main trade-offs emphasized in Example 1 in the Introduction. Consider a society consisting of N individuals and a set of finite states \mathcal{S} . We start with $s_0 = a$ corresponding to absolutist monarchy, where individual E holds power. More formally, $\mathcal{W}_a = \{C \in \mathcal{C} : E \in C\}$. Suppose that for all $x \in \mathcal{S} \setminus \{a\}$, we have that $\{E\} \notin \mathcal{B}_x$, that is, E does not form a blocking coalition on his own (and thus does not form a winning coalition). Moreover, there exists a state, “democracy,” $d \in \mathcal{S}$ such that $\phi(x) = d$ for all $x \in \mathcal{S} \setminus \{a\}$. In other words, starting with any regime other than absolutist monarchy, we will eventually end up with democracy. Suppose also that there exists $y \in \mathcal{S}$ such that

$$w_y(i) > w_a(i),$$

meaning that all individuals are better off in state y than in absolutist monarchy, a . In fact, the gap between the payoffs in state y and those in a could be arbitrarily large. It is straightforward to verify that Assumptions 1–3 are satisfied in this game.

To understand economic interactions in the most straightforward manner, consider the extensive-form game described in Section 4. It is then clear that for β sufficiently large, E will not accept any reforms away from a . In particular, since, given our specification, the game

will reach state d in a finite number of periods, for any gap between $\max_{x \in \mathcal{S}} \{w_x(E)\}$ and $w_a(E)$, there exists a sufficiently large β that E is better off if the state remains to be a .

This example illustrates the potential (and potentially large) inefficiencies that can arise in games of dynamic collective decision-making and emphasizes that commitment problems are at the heart of these inefficiencies. If the society could collectively commit to stay in some state $y \neq d$, then these inefficiencies could be partially avoided. And yet such a commitment is not possible, since once state y is reached, E is no longer a blocking coalition and the rest of the society wishes to progress towards d .

6.2 Middle Class and Democratization

Let us consider a variation of this game. Suppose again that the initial state is $s_0 = a$, where $\mathcal{W}_a = \{C \in \mathcal{C} : E \in C\}$. To start with, suppose that there is only one other agent, P , representing the poor, and two other states, $d1$, democracy with limited redistribution, and $d2$, democracy with extensive redistribution. Suppose that $\mathcal{W}_{d1} = \mathcal{W}_{d2} = \{C \in \mathcal{C} : P \in C\}$. As before,

$$w_{d2}(E) < w_a(E) < w_{d1}(E),$$

and

$$w_a(P) < w_{d1}(P) < w_{d2}(P),$$

so that P prefers extensive redistribution. Given the fact that $\mathcal{W}_{d1} = \mathcal{W}_{d2} = \{P\}$, once democracy is established, the poor can implement extensive a distribution. Anticipating this, E will resist democratization.

Now imagine that an additional social group emerges, M , representing the middle class, and the middle class is sufficiently numerous so that

$$\mathcal{W}_{d1} = \mathcal{W}_{d2} = \{\{M, P\}, \{E, M, P\}\}.$$

Their preferences are also opposed to extensive redistribution, so

$$w_a(M) < w_{d2}(M) < w_{d1}(M).$$

This implies that once state $d1$ emerges, there no longer exists a winning coalition to force extensive redistribution. Now anticipating this, E will be happy to establish democracy (extend the franchise). Therefore, this example illustrates how the presence of an additional player can have moderating effect (see Acemoglu and Robinson, 2006a, for examples in which the middle class may have played such a role in the process of democratization).

6.3 Voting in Clubs

Let us now return to Example 2. Let the society consists of N individuals, so that $\mathcal{I} = \{1, \dots, N\}$. Following Roberts (1999), suppose that there are N states, of the form $s_k = \{1, \dots, k\}$, $1 \leq k \leq N$. Roberts (1999) imposes the following strict single crossing property:

$$\text{for all } l > k \text{ and } j > i, \quad w_{s_l}(j) - w_{s_k}(j) > w_{s_l}(i) - w_{s_k}(i).$$

He then considers two voting schemes: majority voting within a club (where in club s_k one needs more than $k/2$ votes for a change in club size to be implemented) or median voter voting (where the agreement of individual $(k+1)/2$ if k is odd or $k/2$ and $k/2 + 1$ if k is even are needed). Roberts proves that under either rule there are no cycles, and the same set of stable clubs emerges.

To show how Roberts's model is a special case of the analysis here, let us adopt the following simplifying assumption

$$\text{for any } i \in \mathcal{I} \text{ and } k \neq l, \quad w_{s_k}(i) \neq w_{s_l}(i). \quad (9)$$

Though not necessary in this case, this assumption simplifies the analysis, in particular, avoiding certain complications that arise when k is even. Majority and median voting rules imply the following structure of winning coalitions,

$$\mathcal{W}_{s_k}^{maj} = \{\mathcal{C} : |\mathcal{C} \cap s_k| > k/2\}$$

and

$$\mathcal{W}_{s_k}^{med} = \begin{cases} \{\mathcal{C} : (k+1)/2 \in \mathcal{C}\} & \text{if } k \text{ is odd;} \\ \{\mathcal{C} : \{k/2, k/2 + 1\} \subset \mathcal{C}\} & \text{if } k \text{ is even.} \end{cases}$$

In addition, let us also refer to a *Modified Roberts model*, in which only odd-sized clubs are allowed. It is straightforward to verify that Roberts's original proof of existence of equilibria apply without any change to this modified model.

The next proposition establishes that the general framework presented in this paper encompasses Roberts's model. This proposition also shows the link between the mapping ϕ that satisfies Axioms 1–3 and Roberts's *Markov Voting Equilibrium* when attention is restricted to odd-sized clubs. We first provide a definition of Markov Voting Equilibrium concept introduced by Roberts for completeness.

Definition 7 (Markov Voting Equilibrium) *A transition rule $y^*(\cdot)$ is a mapping that corresponds a probability distribution (a lottery over the next states) to each state s_k . Define continuation value of individual i if the current state is s_k and transition rule is $y(\cdot)$ by $V_i(s_k, y(\cdot))$.*

For each state s_k consider set $Y(s_k)$ such that for any $y \in Y(s_k)$ and any state z , the number of players among players $1, \dots, k$ such that $V_i(y, y^*(\cdot)) > V_i(z, y^*(\cdot))$ is at least as large as the number of those with $V_i(y, y^*(\cdot)) < V_i(z, y^*(\cdot))$. Transition rule $y^*(\cdot)$ is a Markov Voting Equilibrium if the support of $y^*(s_k)$ is a subset of $Y^*(s_k)$ for any state s_k .

Note that according to this definition, a majority that strictly prefers a transition is not necessary for a transition to take place, which contrasts with the approach taken in our paper. Nevertheless, the following proposition shows that Roberts's (1999) model and results follow directly from the general framework developed in this paper.

Proposition 1 *1. Assumptions 1 and 2 are satisfied in the original Roberts model and in the Modified Roberts model.*

2. In the Modified Roberts model, Assumption 3 is satisfied and there exists a unique mapping ϕ that satisfies Axioms 1–3.

3. Suppose that β is close to 1 and that a pure strategy Markov Voting Equilibrium exists in the Modified Roberts model. Then, a steady state reached starting with club s_0 in this equilibrium coincides with the dynamically stable state $\phi(s_0)$.

Proof. See Appendix A. ■

6.4 Stable Voting Rules and Constitutions

Let us now return to the question of self-stable coalitions posed in Barbera and Jackson (2005) and discussed in Example 3 in the Introduction. The society takes the form of $\mathcal{I} = \{1, \dots, N\}$, and each state now directly corresponds to a “constitution” represented by a pair (a, b) , where a and b are natural numbers between 1 and N . In Barbera and Jackson's interpretation, a votes are needed to implement a change in some policy variable away from status quo, while b votes are needed to change the current state. Barbera and Jackson consider cases both with $a, b \leq N/2$ and with $a, b > N/2$ (though they note that the former could lead to non-existence of equilibria). Let us simplify the discussion by focusing on the case where $a, b > N/2$. States with $a = b$ correspond to *voting rules*, and those with $a < b$ correspond to *constitutions*, where modifying the current state is more difficult than changing the policy.

Players differ in their ex-ante probability of favoring the change; assume that this probability for player i is $p(i)$, which induces preferences over states $s = (a, b)$. Without loss of generality, assume that $p(i)$ is nondecreasing in i . Barbera and Jackson show that this utility, $w_{(a,b)}(i)$, is “single-set-peaked,” in that if there is more than one peak, these must be two neighboring

peaks and that $w_{(a,b)}(i)$ satisfies a single-crossing condition. Moreover, it can be verified that generically (in terms of perturbations of p 's), there is only one peak and the single-crossing condition holds strictly. To simplify the exposition let us suppose that there is a single peak and that the strict single-crossing property holds.

In Barbera and Jackson, voting rules and constitutions are assumed to have only one-step dynamics (i.e., any deviation ends the game). Hence, stable voting rules and constitutions, as defined in Barbera and Jackson, correspond to myopically, rather than dynamically stable states in our setup. Consequently, it is possible that a rule is unstable in the sense of Barbera and Jackson, but players will not deviate from it if they play an infinite-horizon game (the converse, however, is true: a stable point in the sense of Barbera and Jackson is necessarily stable point in this game). The following result can be established.

Proposition 2 *In the modified version of the model of Barbera and Jackson described above:*

1. *Assumptions 1 and 2 are satisfied.*
2. *There exist mappings ϕ_v for the case of voting rules ($a = b$) and ϕ_c for the case of constitutions ($a \leq b$) that satisfy Axioms 1–3.*
3. *Any stable voting rule a (in the sense of Barbera and Jackson) satisfies $\phi_v(a) = a$.*
4. *Any stable constitution (a, b) (in the sense of Barbera and Jackson) satisfies $\phi_c((a, b)) = (a, b)$. Moreover, any pair (a, b) such that $\phi_c((a, b)) = (a, b)$ is a stable constitution.*

Proof. See Appendix A. ■

Part 4 of Proposition 2 establishes that in the case of constitutions, unlike the case of voting rules, not only myopically stable states correspond to stable constitutions, but any dynamically stable state corresponds to one. This is not a coincidence, but happens because if the players want to switch the constitution, they may always choose the new constitution so that it requires unanimity to be modified. This feature prevents any modification of the new constitution, and serve as a commitment device that is lack in the case of voting rules. However, this effect is not specific to constitutions. Suppose that we slightly altered the setup augmented any voting rule a by adding a state that yields the same payoff as a but cannot be changed (or, alternatively, requires unanimous voting to be changed). In that case, players would always be able to switch to this “additional” state which is impossible to move away from, and this would serve as a commitment device.

6.5 Coalition Formation in Nondemocracies

Let us next turn to the game of dynamic coalition formation we first studied in Acemoglu, Egorov and Sonin (2008). Assume that the set of states \mathcal{S} coincides with the set of coalitions \mathcal{C} . This means that members of the coalition (potentially, both insiders and outsiders) determine the composition of the club in the next period. Assume furthermore that each agent $i \in \mathcal{I}$ is assigned a positive number γ_i which we interpret as “political influence,” and for any coalition $X \in \mathcal{C}$ let

$$\gamma_X = \sum_{j \in X} \gamma_j.$$

Let payoffs be given by

$$w_X(i) = \begin{cases} \gamma_i/\gamma_X & \text{if } i \in X \\ 0 & \text{if } i \notin X \end{cases} \quad (10)$$

for any $i \in \mathcal{I}$ and any $X \in \mathcal{C} \equiv \mathcal{S}$. This is a special case of the payoff structure in Acemoglu, Egorov and Sonin (2008), where we allowed for any payoff function satisfying three properties: if $i \in X$ and $i \notin Y$, then $w_X(i) > w_Y(i)$, if $i \in X$ and $i \in Y$, then $w_i(X) > w_i(Y)$ if and only if $\gamma_i/\gamma_X > \gamma_i/\gamma_Y$, and if $i \notin X$ and $i \notin Y$, then $w_i(X) = w_i(Y)$. The restriction to (10) here is just for simplicity. Also, take any $\alpha \in [1/2, 1)$ as a measure of the extent of supermajority requirement. Define the winning coalitions as

$$\mathcal{W}_X = \left\{ Y : \sum_{j \in Y \cap X} \gamma_j > \alpha \sum_{j \in X} \gamma_j \right\}. \quad (11)$$

Evidently, this corresponds to weighted α -majority voting among members of incumbent coalition X (with $\alpha = 1/2$ corresponding to simple majority). Finally, take π_X to be any permutation of agenda-setters in \mathcal{I} .

It is straightforward to verify that Assumption 1 is satisfied for the set of winning coalitions given by (11). Now suppose that the following simple *genericity* assumption holds:

$$\gamma_X = \gamma_Y \text{ only if } X = Y. \quad (12)$$

Then Assumption 2(a) is also satisfied (no state may be strictly preferred to the one with the least total power). Assumption 2(b) is satisfied whenever the sequence s_1, \dots, s_l has $\gamma_{s_1} > \gamma_{s_l}$. So, while we cannot formally apply Theorems 1, 2, 1', 2', these Theorems are still valid (provided that protocol $\pi_s(k)$ takes the form $\pi_s(k) = X_k$, where X_k is a fixed sequence of all coalitions with strictly decreasing total power. It is also easy to check that Assumption 3 holds.

In Acemoglu, Egorov, and Sonin (2008), we interpreted this game as one of “eliminations” from ruling coalitions in nondemocracies, so that once a particular individual was eliminated, he could no longer be part of future ruling coalitions (either because he is “killed,” permanently

exiled, or is permanently excluded from politics via other means). Moreover, we assumed that payoffs were realized at the end of the game. The results of that model can be represented as a special case of our framework here by using the generalization in Section 5. In particular, suppose that transition $X \rightsquigarrow Y$ is feasible if and only if $Y \subset X$. Clearly, this represents the structure of transitions in the paper and the relation \rightsquigarrow defined in this way satisfies Assumption 5. In addition, the framework developed here also enables us to generalize the results in Acemoglu, Egorov and Sonin (2008) by allowing any transitions to be feasible.

Proposition 3 *Consider the environment in Acemoglu, Egorov, and Sonin (2008). Suppose that the genericity assumption (12) holds.*

1. *Assumptions 1, 2', 3', and 5 are satisfied (provided that $X \rightsquigarrow Y$ is feasible if and only if $Y \subset X$).*
2. *There exists a unique outcome mapping ϕ_{elim} that satisfies Axioms 1'-3'. This mapping yields the same equilibrium (dynamically stable) states as in Acemoglu, Egorov, and Sonin (2008).*
3. *Consider an extended version of this environment where any transition is possible (i.e., $X \rightsquigarrow Y$ is feasible for any $X, Y \in \mathcal{C}$). In this extended version, Assumptions 1, 2 and 3 are satisfied, and there exists a unique outcome mapping ϕ that satisfies Axioms 1-3. This mapping may not yield the same dynamically stable states as in Acemoglu, Egorov, and Sonin (2008).*

Proof. See Appendix A. ■

The next example illustrates both the reasoning of dynamic coalition formation in nondemocracies and also how this proposition generalizes Acemoglu, Egorov, and Sonin's (2008) results by allowing previously-eliminated players being brought back.

Example 5 Start with the case where $\mathcal{I} = \{A, B, C\}$ with $\gamma_A = 3$, $\gamma_B = 4$, $\gamma_C = 5$ and let $\alpha = 1/2$. Evidently, states, or, equivalently, coalitions $\{A\}, \{B\}, \{C\}$ are stable, while $\{A, B\}, \{B, C\}, \{A, C\}$ are not (leading to $\{B\}, \{C\}, \{C\}$, respectively). As a result, coalition $\{A, B, C\}$ is stable: (1) the elimination of any two players can be easily blocked by these two players; (2) the elimination of one player would then lead to the elimination of the weaker of the remaining two players, and therefore it will also be blocked. This example thus illustrates the fundamental insight discussed after Theorem 1 in the clearest possible fashion: the coalition $\{A, B, C\}$ is made stable by the instability of the subcoalitions $\{A, B\}, \{B, C\}$ and

$\{A, C\}$ —because players realize that any deviation from $\{A, B, C\}$ will lead to a further round of elimination.

The example in the previous paragraph has the simple feature that it is never optimal to include more players to a stable coalition. For example, if the society \mathcal{I} included an additional player D with $\gamma_D = 10$, coalition $\{A, B, C, D\}$ would be unstable. However, if, instead, $\mathcal{I} = \{A, B, C, D, E\}$, the inclusion of player E with $\gamma_E = 20$ would make coalition $\{A, B, C, D, E\}$ stable. And yet, even in this case, $\{A, B, C\}$ would not like to include the additional players $\{D, E\}$, since this would strictly reduce the payoffs of each of A, B , and C .

Nevertheless, we can also construct examples in which dynamically stable states are formed by including players into existing coalitions and relatedly, the set of dynamically stable coalitions depends on whether inclusions, as well as eliminations, are all out. For example, suppose that $\mathcal{I} = \{A, B, C, Z\}$ with the same players A, B, C and $\gamma_Z = 9/2$, then $\phi(\{A, B, C\}) = \{A, B, Z\}$, whereas $\phi_{elim}(\{A, B, C\}) = \{A, B, C\}$. This means that mappings ϕ and ϕ_{elim} may, and generally will, be different.

Another generalization is also of interest. Let us return to one of the examples used by Acemoglu, Egorov and Sonin (2008) to motivate this model, that of Soviet Politburo eliminations. The history of the Soviet Politburo includes cases in which individuals were removed from the Politburo but were left alive as well as the majority of the cases in which those removed from the Politburo were executed. This situation can be captured by considering a pair of subsets of \mathcal{I} , (X, Y) , with $X \subset Y$, where Y denotes the set of politicians who are alive and X corresponds to the set of those at power. Any politician who is alive can be brought back into the ruling coalition, but “resurrections” are not possible. Suppose also that execution is costly. It can be verified that this modified game also satisfies Assumptions 1, 2, 3, and 5 and that the results of Proposition 3 still hold. In this case, we can construct examples where eliminations take place without executions and also examples with executions. For instance, starting with the example of $\{A, B, C, D\}$ above, player D will be excluded from the ruling coalition but will not be executed, since there is never any danger that he will be brought back. However, if there are six players with powers 100, 101, 103, 107, 115, and 131, then provided that the cost of execution is not too high, $\{100, 101, 131\}$ will form the ruling coalition and will execute 103, 107, and 115. This is because if any one of these three players survived, 100 and 101 would use him to replace 131, thus making $\{100, 101, 131\}$ unstable to start with.

6.6 Coalition Formation in Democracy

We next briefly discuss how similar issues are present in the context of coalition formation in democratic situations, for example, coalition formation the context of legislative bargaining (see, for example, Baron and Ferejohn, 1986, Austen-Smith and Banks, 1988, Baron, 1991, Jackson and Moselle, 2002, Norman, 2002, for models of legislative bargaining).

Suppose that there are three parties in the parliament, 1, 2, 3, and any two of them would be sufficient to form a government. Suppose that party 1 has more seats than party 2, which in turn has more seats than party 3. The initial state is \emptyset , and all coalitions are possible states. Since any two parties are sufficient to form a government, we have that $\mathcal{W}_{\emptyset} = \{\{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$. First, suppose that all governments are equally strong and a party with a greater share of seats in the parliament will be more influential in the coalition government. Consequently, $w_{\{1,2\}}(3) = 0 < w_{\{1,2,3\}}(3) < w_{\{1,3\}}(3) < w_{\{2,3\}}(3)$; other payoffs are defined similarly. In this case, it can be verified that $\phi(\emptyset) = \{2, 3\}$: indeed, neither party 2 nor party 3 wishes to form a coalition with party 1, because party 1's influence in the coalition government would be too strong. The equilibrium in this example then coincides with the minimum winning coalition.

However, as emphasized in the Introduction, the dynamics of coalition formation do not necessarily lead to minimum winning coalitions. To illustrate this, suppose that governments that have a greater number of seats in the parliament are stronger, so that $w_{\{1,3\}}(2) = 0 < w_{\{1,2,3\}}(2) < w_{\{2,3\}}(2) < w_{\{1,2\}}(2)$. That is, party 2 receives a higher payoff even though it is a junior partner in the coalition $\{1, 2\}$, because this coalition is sufficiently powerful. We might then expect that $\{1, 2\}$ may indeed arise as the equilibrium coalition, that is, $\phi(\emptyset) = \{1, 2\}$. Nevertheless, whether this will be the case depends on how the coalition $\{1, 2\}$ will play out. Suppose, for example, that after the coalition $\{1, 2\}$ forms, party 1, by virtue of its greater number of seats, can sideline party 2 and rule by itself. Let us introduced the shorthand symbol “ \mapsto ” to denote this, so that we have $\{1, 2\} \mapsto \{1\}$ (which naturally presumes that $\mathcal{W}_{\{1,2\}} = \{C \in \mathcal{C} : 1 \in C\}$). Similarly, starting from the coalition $\{2, 3\}$, party 2 can also do the same starting from the coalition, so that $\mathcal{W}_{\{2,3\}} = \{C \in \mathcal{C} : 2 \in C\}$ and $\{2, 3\} \mapsto \{2\}$. However, is also reasonable to suppose that once party 2 starts ruling by itself, then party 1 can regain power by virtue of its greater seat share, that is, $\mathcal{W}_{\{2\}} = \{C \in \mathcal{C} : 1 \in C\}$ and $\{2\} \mapsto \{1\}$. In this case, the analysis in this paper immediately shows that $\phi(\emptyset) = \{2, 3\}$, that is, the coalition $\{2, 3\}$ emerges as the dynamically stable state.

What makes $\{2, 3\}$ dynamically stable in this case is the fact that $\{2\}$ is not dynamically stable itself. This example therefore reiterates the fundamental principle discussed after Theorem

1 in Section 3 and also in the context of coalition formation in nondemocracies in the previous subsection: the instability of states that can be reached from a state s contributes to the stability of state s .

6.7 Concessions in Civil War

Finally, let us briefly consider an application of the ideas in this paper to the analysis of civil wars. This example can also be used to illustrate how similar issues arise in the context of international wars (see Fearon, 1996, 2004, and Powell, 1998, for more complete discussions of issues related to commitment problems in civil and international wars). The idea that the possibility to face war in some subsequent state of the world might affect the willingness of the parties to launch war right away may be traced back to Fearon (1998), where, motivated by the break out of ethnic conflicts in former Yugoslavia, majority and minority group negotiate in order to avoid a civil war. First, the majority makes a take-it-or-leave-it offer, which the weak party can accept or reject (and then fight). If the minority accepts, it becomes even weaker in the next round. Now, the majority might require even more concessions as the power shifted further from the minority to the majority. Anticipating this, the minority might prefer to fight in the first round than accept the offer. In a continuous-time setting, Schwarz and Sonin (2008) demonstrate that a war might be more likely when the minority is risk-averse as it might fear that a concession might lead to the possibility of having to fight a war in even less advantageous circumstances.

Formally, suppose that a government, G , is engaged in a civil war with a rebel group, R . The civil war state is denoted by c . The government can initiate peace and transition to state p , so that $\mathcal{W}_c = \{C \in \mathcal{C} : G \in C\}$. However, using the shorthand “ \mapsto ” introduced in the previous subsection, we now have $p \mapsto r$, where r denotes a state in which the rebel group becomes very strong and dominant in domestic politics. Moreover, $\mathcal{W}_p = \{C \in \mathcal{C} : R \in C\}$, and naturally, $w_r(R) > w_p(R)$. If $w_r(G) < w_c(G)$, there will be no peace and $\phi(c) = c$ despite the fact that we may also have $w_p(G) > w_c(G)$. This illustrates the same forces as in the example on resistance to efficiency-enhancing reforms or to beneficial institutional changes discussed above.

As an interesting modification, suppose next that the rebel group R can first disarm partially, in particular, $c \mapsto d$, where d denotes the state of partial disarmament. Moreover, $d \mapsto dp$, where the state dp involves peace with the rebels that have partially disarmed. Suppose that $\mathcal{W}_{dp} = \{\{G, R\}\}$, meaning that once they have partially disarmed, the rebels can no longer become dominant in domestic politics. In this case, provided that $w_{dp}(G) > w_d(G)$, we have $\phi(c) = dp$. Therefore, the ability of the rebel group to make a concession changes the set of

dynamically stable states. This example therefore shows how the role of concessions can also be introduced into this framework in a natural way.

7 Conclusion

A central feature of collective decision-making in many social groups, such as political coalitions, international unions, or private clubs, is that the rules that govern regulations, procedures for future decision-making, and inclusion and exclusion of members are made by the current members and under the current regulations. This feature implies that dynamic collective decisions must recognize the implications of current decisions on future choices. For example, current constitutional change must recognize how the new constitution will open the way for further changes in laws and regulations and how these further changes might affect the long-run payoffs of different players.

In this paper, we develop a general framework for a systematic study of this class of problems. We provide both an axiomatic and a noncooperative characterization of stable states and show that, under relatively mild assumptions, they exist and are (essentially) unique. We show that the set of dynamically stable states can be computed recursively. This recursive characterization highlights that a particular state s is dynamically stable if there does not exist another dynamically stable state that makes a winning coalition (within s) better off. This characterization therefore highlights two important economic features of dynamic decision-making in situations where rules are set by currently politically powerful groups:

1. It is the instability of potential states that have enough supporters that makes a particular state dynamically stable. In particular, a state can be dynamically stable even if there is a politically powerful (winning) coalition that prefers an alternative state. This is because the alternative state itself might not be dynamically stable and a transition to this state would be followed by another transition leading ultimately to a state that makes some of the members of the initial (winning) coalition worse off.
2. The same reasoning implies that Pareto inefficient states may arise as dynamically stable states. In particular, reforms towards Pareto preferred states may not be undertaken because these states themselves might be unstable and would lead to further changes in payoffs.

We show that our framework is general enough to nest various different models that have been used in the literature to analyze specific problems in which current collective decisions affect

future decision-making procedures. These include models of inefficient inertia (lack of reform) because of fear of changes in the future balance of political power, models of institutional change and enfranchisement (such as Acemoglu and Robinson, 2001, 2006a; Lizzeri and Persico, 2004; Jack and Lagunoff, 2006), models of voting in clubs (such as Roberts, 1999; Barbera, Maschler, and Shalev, 2001), models of the stability of constitutions (such as Barbera in Jackson, 2004), and models of coalition formation in democracies or nondemocracies. In these cases and in a number of others, we show how the exact model previously studied in the literature (or a generalization or a slight variant thereof) is a special case of our framework and how this allows us to derive the main economic insights in a straightforward manner.

Although our framework is fairly general, our analysis still relies on a number of important assumptions. Some of those are necessary for our general approach (for example, a minimum amount of acyclicity is essential). Others are adopted for convenience and can be relaxed, though often at the cost of further complication. Among possible extensions, we believe that most interesting would be to introduce stochastic elements, so that the set of feasible transitions or the distribution of powers change over time, or to include additional state variables, such as capital, so that some subcomponent of the state variables have autonomous dynamics. Such extensions would allow us to incorporate an even larger set of dynamic political games within this framework. We view the analysis of such dynamics as an interesting area for future research.

Appendix A

We start with a lemma about the structure MPE in voting-type games, which will be used in the proof of Theorem 2. This lemma shows that there always exists a pure strategy MPE in which each individual votes for the outcome that he or she strictly prefers and that in any (mixed or pure strategy) equilibrium the outcome that is preferred by a majority will be implemented.

Lemma 1 *Consider an extensive-form game G with perfect information and with N stages which takes the following form. In each stage k , one player i_k (this player may be the same for different k 's) makes action a_k which may be y or n . There exist two payoff vectors, \bar{y} and \bar{n} , such that for each terminal node reached after sequence of actions (a_1, \dots, a_N) the payoff vector $v(a_1, \dots, a_N)$ is either \bar{y} or \bar{n} . In addition, assume that if for some vector of actions (a_1, \dots, a_N) and for some k , if $v(a_k = n, a_{-k}) = \bar{y}$ (where a_{-k} is the set of all actions other than k), then $v(a_k = y, a_{-k}) = \bar{y}$. Then:*

(i) *There exists a Markov Perfect Equilibrium in pure strategies where action $a_k = y$ if $v_{i_k}(\bar{y}) > v_{i_k}(\bar{n})$ and $a_k = n$ if $v_{i_k}(\bar{y}) < v_{i_k}(\bar{n})$.*

(ii) *Suppose that the set of players $Y = \{i : v_i(\bar{y}) > v_i(\bar{n})\}$ is large enough, in the sense that $v(a_1, \dots, a_N) = \bar{y}$ whenever $a_k = y$ for all $i_k \in Y$. Then in any Subgame Perfect Nash equilibrium, the equilibrium payoff vector will be \bar{y} with probability 1. Similarly, if the set of players $N = \{i : v_i(\bar{y}) < v_i(\bar{n})\}$ is large enough, so that $v(a_1, \dots, a_N) = \bar{n}$ whenever $a_k = n$ for all $i_k \in N$, then in any Markov Perfect Equilibrium payoff vector will be \bar{n} with probability 1.*

(iii) *Suppose that the first N stages of a finite or infinite extensive-form game G with perfect information satisfy the requirements above, except that instead of payments at terminal nodes taking two values only, we have that there are exactly two classes of isomorphic subgames, \mathcal{Y} and \mathcal{N} . Take any Markov Perfect equilibrium σ and let $Y = \{i : v_i(\mathcal{Y}) > v_i(\mathcal{N})\}$ and let $N = \{i : v_i(\mathcal{Y}) < v_i(\mathcal{N})\}$, where $v_i(\mathcal{Y})$ and $v_i(\mathcal{N})$ are continuation payoffs of player i (in a MPE, these are well-defined). If we have $v(a_1, \dots, a_N) = \bar{y}$ whenever $a_k = y$ for all $i_k \in Y$, then continuation game reached is \mathcal{Y} , and the expected utility players receive in this MPE is given by $v(\mathcal{Y})$. Conversely, if $v(a_1, \dots, a_N) = \bar{n}$ whenever $a_k = n$ for all $i_k \in N$, then continuation game reached is \mathcal{N} , and the expected utility players receive in this MPE is given by $v(\mathcal{N})$.*

Proof of Lemma 1 (Part 1) We need to show that for the profile of strategies in which $a_k = y$ if $v_{i_k}(\bar{y}) > v_{i_k}(\bar{n})$ and $a_k = n$ if $v_{i_k}(\bar{y}) < v_{i_k}(\bar{n})$ (and a_k is either y or n if $v_{i_k}(\bar{y}) = v_{i_k}(\bar{n})$), there is no profitable deviation for any player at any stage. Consider player i_k for whom $v_{i_k}(\bar{y}) > v_{i_k}(\bar{n})$; he plays $a_k = y$. If he switches to $a'_k = n$, this would not change the action of any of the subsequent voters, and therefore this either would not change the outcome of the

voting (i.e., the payoff vector) or will change it from \bar{y} to \bar{n} . In both cases this deviation is not profitable. Similarly, for player i_k with $v_{i_k}(\bar{y}) < v_{i_k}(\bar{n})$, deviation from $a_k = n$ to $a'_k = y$ may only change the payoff vector from \bar{n} to \bar{y} only, which is not profitable for such player. Finally, if for player i_k , $v_{i_k}(\bar{y}) = v_{i_k}(\bar{n})$, then any outcome yields the same utility to him, and he does not have a profitable deviation regardless of the action that he plays. This completes the proof.

(Part 2) We prove this by induction by the number of stages k . Base: take $k = 1$. Suppose that set Y is large enough, so that if the single player i_1 chooses action $a_1 = y$, then the payoff vector is \bar{y} . To obtain a contradiction, suppose that in a SPNE the equilibrium payoff vector may be different from \bar{y} with a positive probability, in which case the payoff vector is \bar{n} . But then player i_1 is better off if he chose action $a_1 = y$ with probability 1, since he would then receive $v_{i_1}(\bar{y})$, which cannot be the case in an equilibrium. We can similarly consider the case where set N is large enough. We have thus proved the base of induction.

Step: suppose that we have proved the result for all $l \leq k - 1$; consider game with k stages. Suppose that set Y is large enough (the case where set N is large enough may be treated similarly). Consider two cases. Suppose first that $v_{i_1}(\bar{y}) > v_{i_1}(\bar{n})$. If player i_1 in stage 1 takes action $a_1 = y$, then in the subgame which starts at stage 2 we would have that if all players for whom $v_{i_j}(\bar{y}) > v_{i_j}(\bar{n})$ for all $2 \leq j \leq k$ will choose $a_j = y$, then the payoff vector will be \bar{y} . By induction, any SPNE in this subgame will lead to \bar{y} with probability 1. Now if for some SPNE of the entire game the payoff vector is \bar{n} with a positive probability, then player i_1 may ensure that the payoff vector is \bar{y} with probability 1 by undertaking action $a_1 = y$, so he has a profitable deviation. Therefore, in this case, the payoff vector is \bar{y} with probability 1.

Now suppose that $v_{i_1}(\bar{y}) \leq v_{i_1}(\bar{n})$. Then, by assumption, if in the subgame which starts at stage 2 all players for whom $v_{i_j}(\bar{y}) > v_{i_j}(\bar{n})$ for all $2 \leq j \leq k$ will choose $a_j = y$, then the payoff vector will be \bar{y} . By induction, for any SPNE in any subgame which starts at stage 1, the payoff vector is \bar{y} with probability 1. But this implies that the same holds for the entire game. This completes the induction step for the case where Y is large enough. The case where N is large enough is analogous, and this completes the proof of Part 2.

(Part 3) This immediately follows from Part 2, since a MPE induces a SPNE in the reduced game of first k stages with payoffs given by continuation payoffs of the original game. ■

Proof of Theorem 2 (Part 1) First, we require that β satisfies the following conditions: if $w_{s'}(i) < w_s(i)$, then $\beta^{|\mathcal{S}|} > w_{s'}(i)/w_s(i)$, and $(1 - \beta)/\beta < (w_s(i) - w_{s'}(i)) / \max_{q \in \mathcal{S}} w_q(i)$ for any i, s, s' . Since this gives us a finite (or empty, in degenerate cases) number of inequalities, there exists $\beta_0 < 1$ such that for all $\beta > \beta_0$ these condition is satisfied.

We construct a MPE of the game with the following property: for each period $t \geq 1$, $s_t = \phi(s_{t-1})$. We introduce the following notation: for $i \in \mathcal{I}$ and $s, q \in \mathcal{S}$, let

$$v_s^q(i) = \left\{ \begin{array}{ll} (1 - \beta) w_s(i) & \text{if } s = q \\ 0 & \text{if } s \neq q \end{array} \right\} + \left\{ \begin{array}{ll} \beta w_{\phi(q)}(i) & \text{if } \phi(q) = q \\ \beta^2 w_{\phi(q)}(i) & \text{if } \phi(q) \neq q \end{array} \right\}. \quad (\text{A1})$$

In the equilibrium we construct below, this will equal the continuation payoff of a player i if the current state is s and the accepted proposal is q if $q \neq s$, and if no proposal is accepted if $q = s$. We drop time index for notational convenience, as we are constructing a MPE.

For each $s \in \mathcal{S}$, take $K_s \geq |\mathcal{S}| - 1$. Take $\pi_s(\cdot)$ such that $\pi_s(K_s) = \phi(s)$ if $\phi(s) \neq s$; otherwise, take $\pi_s(k)$ arbitrarily, making sure that Assumption 4 is satisfied. We construct strategies as follows: if $\phi(s) \neq s$, in the last voting (stage K_s), player i votes for $P_{K(s)} = \phi(s)$ (says *yes*) if and only if $v_s^{\phi(s)}(i) > v_s^s(i)$. Otherwise, player i votes for proposal P_k if and only if $v_s^{P_k}(i) > v_s^{\phi(s)}(i)$, and if $\pi_s(k) \in \mathcal{I}$ for some k , this player chooses proposal P_k arbitrarily. The strategies that we construct in this way are Markovian.

Let us show that if these strategies are played, then there is no transition if $\phi(s) = s$ and there is transition to $\phi(s)$ if $\phi(s) \neq s$. If $\phi(s) \neq s$, the set of players for whom $v_s^{\phi(s)}(i) > v_s^s(i)$ is a winning coalition (in s) by Axiom 1: indeed, for a winning coalition of players $w_{\phi(s)}(i) > w_s(i)$, which implies $\beta w_{\phi(s)}(i) > w_s(i)$ since $\beta > \beta_0$, and therefore for such players

$$v_s^{\phi(s)}(i) = \beta w_{\phi(s)}(i) > w_s(i) + \beta^2 w_{\phi(s)}(i) = v_s^s(i).$$

Let us prove that the set of players for whom $v_s^{P_k}(i) > v_s^{\phi(s)}(i)$ does not form a winning coalition in s . Indeed, since $P_k \neq s$ because there is voting, this inequality implies

$$\beta w_{\phi(P_k)}(i) \geq v_s^{P_k}(i) > v_s^{\phi(s)}(i) \geq \beta w_{\phi(s)}(i),$$

because $\phi(\phi(s)) = \phi(s)$, and therefore $w_{\phi(P_k)}(i) > w_{\phi(s)}(i)$. If such players formed a winning coalition, we would have $\phi(P_k) \succ_s \phi(s)$, which, given that $\phi(s) \succ_s s$, implies that $\phi(P_k) \succ_s s$ by Assumption 2(b)*, but existence of $\phi(P_k)$ such that $\phi(P_k) \succ_s \phi(s)$, $\phi(P_k) \succ_s s$, and $\phi(P_k)$ is stable contradicts Axiom 2. This proves that the set of players with $v_s^{P_k}(i) > v_s^{\phi(s)}(i)$ does not form a winning coalition in s , and therefore, no proposal is accepted, except for the last proposal $P_{K(s)} = \phi(s)$ if $\phi(s) \neq s$.

Now, let us check that if these strategies are played, continuation payoffs after acceptance of proposal q are indeed given by (A1). If proposal $q \neq s$ is accepted, then there is an immediate transition, and there is another transition next period in case $\phi(q) \neq q$. If no proposal is accepted, so $q = s$, then there is no transition in the current period, and each player i receives $w_s(i)$, and there is transition next period if $\phi(s) = \phi(q) \neq q = s$. In either case, the continuation payoffs are given by (A1).

Finally, let us check that no player has an incentive to deviate. For an agenda-setter, this holds because no proposal that an agenda-setter makes is accepted. For a voter, this follows from Lemma 1(a): the continuation strategies are Markovian, and therefore each voting constitutes a finite game with two possible outcomes. Hence, it is always a best response for a voter to vote for the option that he weakly prefers. If $\phi(s) \neq s$, then in the last voting, each player i compares continuation payoff $v_s^{\phi(s)}(i)$ if the proposal is accepted and $v_s^s(i)$ if it is rejected; in all other votings, player i receives $v_s^{P_k}(i)$ if proposal P_k is accepted and $v_s^{\phi(s)}(i)$ if it is rejected (because $\phi(s)$ will be eventually accepted if $\phi(s) \neq s$ and no proposal will be accepted, in which case each player will receive $v_s^{\phi(s)}(i) = v_s^s(i)$ if $\phi(s) = s$). Therefore, no voter can profitably deviate from the strategy specified. This completes the proof that the strategy profile constructed forms a MPE.

(Part 2) Take any set of sequences $\{\pi_s(\cdot)\}$ and any MPE. For any state s , the proposal q which is accepted on the equilibrium path is well-defined (with $q = s$ if all proposals are rejected) since we consider equilibria in pure strategies; denote $\chi(s) = q$. The mapping $\chi : \mathcal{S} \rightarrow \mathcal{S}$, obtained in this way, has no cycles, in the sense that if $\chi(s) \neq s$ then for any $n > 1$, iteration $\chi^n(s) \neq s$. Indeed, assume, to obtain a contradiction, that there is such n . Denote by $J_s \subset \{1, \dots, K_s\}$ the set of votings in state s where proposal P_k made on equilibrium path is accepted (this proposal and whether it is accepted do not depend on the play before current stage k , since strategies are Markovian); evidently, the first voting in J_s leads to $\chi(s)$ (this is what happens in equilibrium). Consider two cases. If all votings in J_s lead to cycles, then consider the last voting. Each player knows that if proposal P is accepted, he will receive zero utility, while if it is rejected, he will receive $(1 - \beta)w_s(i) > 0$. From Lemma 1(c), we can then conclude that P may not be accepted in a MPE, which is a contradiction. In the other case, where not all votings in J_s lead to cycles; denote the votings that do not lead to cycles by $J'_s \subset J_s$. Consider the last voting in J_s that precedes the first voting in J'_s . In this voting, each player knows that accepting leads to zero utility, while rejecting leads to a positive payoff. Therefore, proposal P may not be accepted, which again leads to a contradiction.

The absence of cycles implies that sequence $\chi^n(s)$ stabilizes after no more than $|\mathcal{S}| - 1$ transitions; denote $\psi(s) = \chi^{|\mathcal{S}|}(s)$, and let

$$m(s) = \min \{n \in \mathbb{N} \cup \{0\} : \chi^n(s) = \psi(s)\},$$

where we defined $\chi^0(s) = s$. Evidently, $0 \leq m(s) \leq |\mathcal{S}| - 1$, and $m(s) = 0$ if and only if $\psi(s) = \chi(s) = s$; in addition,

$$\psi(\psi(s)) = \chi(\psi(s)) = \psi(\chi(s)) = \psi(s)$$

for any state s , where the first equality follows from $\chi(\psi(s)) = \psi(s)$ and definition of mapping ψ . Let us denote

$$\bar{v}_s^q(i) = \left\{ \begin{array}{ll} (1 - \beta) w_s(i) & \text{if } s = q \\ 0 & \text{if } s \neq q \end{array} \right\} + \beta^{m(q)+1} w_{\psi(q)}(i); \quad (\text{A2})$$

from the discussion above it follows that $\bar{v}_s^A(i)$ gives the continuation payoff of player i if in state s alternative A is implemented, and after that equilibrium play follows. We now show that mapping $\psi(s)$ satisfies Axioms 1–3, and then we will prove that $\chi(s) = \psi(s)$, i.e., once transition to $\chi(s)$ is completed, there are no more transitions, and $\chi(s)$ is the dynamically stable state reached with zero or one transition.

For each state s consider, as before, the set of votings J where corresponding proposals P_k , $k \in J$, is accepted; let $J = \{k_1, \dots, k_{|J|}\}$, where $k_j < k_l$ for $j < l$ (we drop index s for convenience), and suppose that $J \neq \emptyset$ (this implies $\chi(s) \neq s$ and $m(s) \geq 1$). In equilibrium, proposal P_{k_1} is accepted, so $\psi(P_{k_1}) = \psi(s)$ and $m(P_{k_1}) = m(s) - 1$. Since each of proposals P_{k_l} for $1 \leq l \leq |J|$ is accepted in this equilibrium, we must have, again by Lemma 1(c), that $\psi(P_{k_l}) \succeq \psi(P_{k_{l+1}})$ (only players who weakly prefer $\psi(P_{k_l})$ to $\psi(P_{k_{l+1}})$ may vote for acceptance, for if

$$w_{\psi(P_{k_l})}(i) < w_{\psi(P_{k_{l+1}})}(i),$$

then $\bar{v}_s^{P_{k_l}}(i) < \bar{v}_s^{P_{k_{l+1}}}(i)$ since $\beta \geq \beta_0$). In addition, we have $\psi(P_{k_{|J|}}) \succeq \psi(P_{k_1})$: indeed, if $P_{k_{|J|}}$ is accepted, each player i will receive

$$\bar{v}_s^{P_{k_{|J|}}}(i) = \beta^{m(P_{k_{|J|}})+1} w_{\psi(P_{k_{|J|}})}(i),$$

while if it is rejected, each player will receive

$$\bar{v}_s^s = (1 - \beta) w_s(i) + \beta^{m(s)+1} w_{\psi(P_{k_1})}(i).$$

If

$$w_{\psi(P_{k_{|J|}})}(i) < w_{\psi(P_{k_1})}(i),$$

then

$$\beta^{m(P_{k_{|J|}})+1} w_{\psi(P_{k_{|J|}})}(i) < \beta^{m(s)+1} w_{\psi(P_{k_1})}(i),$$

and hence

$$\bar{v}_s^{P_{k_{|J|}}}(i) < \bar{v}_s^s.$$

Since $P_{k_{|J|}}$ is accepted, the set of such players does not form a blocking coalition, which implies

$$w_{\psi(P_{k_{|J|}})}(i) \geq w_{\psi(P_{k_1})}(i)$$

for a winning coalition of players, establishing that $\psi(P_{k_{|J|}}) \succeq \psi(P_{k_1})$. Now, since Assumption 2(b)* holds, we have $\psi(P_{k_j}) \sim \psi(P_{k_l})$ for all $1 \leq j < l \leq |J|$. In addition, we prove that $m(P_{k_l}) \leq m(P_{k_{l+1}})$ for all $1 \leq l \leq |J| - 1$. Indeed, if this were not the case, each player would receive a strictly higher payoff if P_{k_l} was rejected at stage k_l , so P_{k_l} could not be accepted in the equilibrium.

In what follows, we establish that mapping $\psi(s)$ satisfies Axioms 1–3 and that there is at most one transition. First, we prove that Axiom 1 holds. Consider set J introduced above and consider stage $k_{|J|}$, i.e., the last stage where acceptance is possible. If $P_{k_{|J|}}$ is accepted, each player receives

$$\bar{v}_s^{P_{k_{|J|}}}(i) = \beta^{m(P_{k_{|J|}})+1} w_{\psi(P_{k_{|J|}})}(i).$$

If $P_{k_{|J|}}$ is rejected, each player receives

$$\bar{v}_s^s = (1 - \beta) w_s(i) + \beta^{m(s)+1} w_{\psi(s)}(i).$$

We have established, however, that $w_{\psi(P_{k_{|J|}})}(i) = w_{\psi(s)}(i)$ and that $m(s) = m(P_{k_1}) + 1 \leq P_{k_{|J|}} + 1$. Since $P_{k_{|J|}}$ is accepted in equilibrium, by Lemma 1(c) we must have that $\bar{v}_s^{P_{k_{|J|}}}(i) \geq \bar{v}_s^s$ for a winning coalition of players (in s). For such players, we have

$$\beta^{m(s)} w_{\psi(s)}(i) \geq \beta^{m(P_{k_{|J|}})+1} w_{\psi(s)}(i) \geq (1 - \beta) w_s(i) + \beta^{m(s)+1} w_{\psi(s)}(i),$$

hence, $\beta^{m(s)} w_{\psi(s)}(i) \geq w_s(i)$, which, given that $\beta > \beta_0$, implies that $w_{\psi(s)}(i) \geq w_s(i)$. However, if $w_{\psi(s)}(i) = w_s(i)$, we would have $\beta^{m(s)} w_{\psi(s)}(i) < w_s(i)$, because $m(s) \geq 1$, since $\psi(s) \neq s$. Consequently, $w_{\psi(s)}(i) > w_s(i)$ for a winning coalition of players, establishing that $\psi(s) \succ s$.

It is straightforward to show that Axiom 3 holds: indeed, we have already showed that for any state s , $\psi(\psi(s)) = \psi(s)$ by definition of mapping ψ . Let us now prove that if $\psi(s) \neq s$, then transition to state $\psi(s)$ takes place in one step, i.e., that $\psi(s) = \chi(s)$, or, equivalently, $m(s) = 1$ whenever $\chi(s) \neq s$. Consider two possibilities. If $\psi(s) = P_{k_j}$ for some $j : 1 \leq j \leq |J|$, then $m(P_{k_l}) = 0$ since Axiom 3 holds. But we proved that $m(P_{k_l})$ is weakly increasing in l , therefore, $m(\chi(s)) = m(P_{k_1}) = 0$, and therefore $m(s) = 1$. The other possibility is that $\psi(s) = P_{k_j}$ does not hold for any j ; this implies, in particular, that $m(P_{k_1}) \geq 1$ and $\psi(s) \neq \chi(s)$. Note that in this case, if for some $k \notin J$ we have $P_k = \psi(s)$ (regardless of whether this happens on or off equilibrium path), then P_k should be accepted. Indeed, take any player i . If $P_k = \psi(s)$ is accepted, this player will receive $\bar{v}_s^{\psi(s)}(i) = \beta w_{\psi(s)}(i)$, while if it is rejected, he will receive

$$\bar{v}_s^{P_{k_l}}(i) = \beta^{m(P_{k_l})+1} w_{\psi(s)}(i) \leq \beta^2 w_{\psi(s)}(i)$$

for some l if $k < k_{|J|}$ and

$$\bar{v}_s^s(i) = (1 - \beta) w_s(i) + \beta^{m(P_{k_1})+1} w_{\psi(s)}(i) \leq (1 - \beta) w_s(i) + \beta^2 w_{\psi(s)}(i)$$

if $k > k_{|J|}$. In the first case, all players prefer to have P_k accepted, while in the second case, each player with $w_{\psi(s)}(i) > w_s(i)$ will have $\beta w_{\psi(s)}(i) > w_s(i)$ since $\beta > \beta_0$, and therefore $\bar{v}_s^{\psi(s)}(i) > \bar{v}_s^s(i)$; since such players form a winning coalition, this implies that $P_k = \psi(s)$ will necessarily be accepted. Since we know that this $k \notin J$, we must conclude that proposal $\psi(s)$ is never considered, which, by Assumption 4, implies that all players become agenda-setters for some k . But then take any player i such that $w_{\psi(s)}(i) > w_s(i)$ and suppose that he is agenda-setter at stage k . He knows that his equilibrium proposal will give him utility $\bar{v}_s^{P_{k_l}}(i)$ for some l if $k < k_{|J|}$ and $\bar{v}_s^s(i)$ if $k > k_{|J|}$. However, proposing $\psi(s)$ will give him a strictly higher utility $\bar{v}_s^{\psi(s)}(i)$, as we showed before. Therefore, he must have a profitable deviation, which cannot be the case in equilibrium. This contradiction shows that the case where $\psi(s) = P_{k_j}$ does not hold for any j is impossible, and this proves that transition to state $\psi(s)$ takes place in one step, so $\psi(s) = \chi(s)$.

Finally, let us show that Axiom 2 holds. Assume, to obtain a contradiction, that this is not the case, i.e., for some state s there exists state z with $\psi(z) = \chi(z) = z$, $z \succ s$ (which implies $z \neq s$), and $z \succ \chi(s) = \psi(s)$ (which implies $\psi(z) \approx \psi(s)$). As before, we can prove that if $P_k = z$ for some k , then proposal P_k must be accepted. Indeed, accepting proposal z will lead to utility $\bar{v}_s^z(i) = \beta w_z(i)$ for any player i , while rejecting it will lead either to $\bar{v}_s^{P_{k_l}}(i) \leq \beta w_{\psi(s)}(i)$ for some l if $J \neq \emptyset$ and $k < k_{|J|}$ or

$$\bar{v}_s^s(i) = (1 - \beta) w_s(i) + \beta^{m(s)+1} w_{\psi(s)}(i) \leq (1 - \beta) w_s(i) + \beta w_{\psi(s)}(i)$$

if $J = \emptyset$ or $k > k_{|J|}$. Consider player i for whom $w_z(i) > w_{\psi(s)}(i)$; such players form a winning coalition. In the first case, we can conclude

$$\bar{v}_s^z(i) = \beta w_z(i) > \beta w_{\psi(s)}(i) \geq \bar{v}_s^{P_{k_l}}(i).$$

In the second case, $\bar{v}_s^z(i) > \bar{v}_s^s(i)$ follows from the fact that

$$\beta w_z(i) \geq (1 - \beta) w_s(i) + \beta w_{\psi(s)}(i),$$

which is a consequence of the assumption that $\beta \geq \beta_0$. Again, by Lemma 1(c) we conclude that proposal $P_k = z$ must be accepted. From this, we can see that z is never proposed, because for any proposals that are accepted (and z is accepted if proposed), as we showed, we would have

$\psi(z) \sim \psi(s)$, which is not the case. By Assumption 4 we obtain that all players become agenda-setters for some k . Again, consider some player i for whom $w_z(i) > w_{\psi(s)}(i)$ and suppose that he is the agenda-setter at some stage k . If he makes his equilibrium proposal, he receives either

$$\bar{v}_s^{P_{k_l}}(i) \leq \beta w_{\psi(s)}(i),$$

where $1 \leq l \leq |J|$, or

$$\bar{v}_s^s(i) = (1 - \beta)w_s(i) + \beta^{m(s)+1}w_{\psi(s)}(i) \leq (1 - \beta)w_s(i) + \beta w_{\psi(s)}(i),$$

while if he makes proposal $P_k = z$, he will receive $\bar{v}_s^z(i) = \beta w_z(i)$. But we already showed that for such player i , the latter utility $\beta w_z(i)$ is strictly higher than both $\beta w_{\psi(s)}(i)$ and $(1 - \beta)w_s(i) + \beta w_{\psi(s)}(i)$. This means that player i has a profitable deviation, which brings us to a contradiction.

We have thus proved that mapping ψ satisfies Axiom 2. This completes the proof of Part 2.

(Part 3) This result immediately follows from Theorem 1 and Part 2 of this Theorem. ■

Proof of Proposition 1 (Part 1) Assumption 1 is satisfied trivially. Let us check that Assumption 2 is satisfied in the original Roberts model (the argument for the Modified Roberts model is identical). For part (a), take any collection of states $\mathcal{Q} \subset \mathcal{S}$; suppose these states are s_{k_1}, \dots, s_{k_l} for some increasing sequence $\{k_j\}_{j=1}^l$. Take state s_{k_1} ; if for any other state there is a blocking coalition that weakly prefers s_{k_1} to this state then we are done, otherwise there is state s_{k_j} such that majority $s_{k_j} \succ_{s_{k_1}}^{maj} s_{k_1}$; without loss of generality suppose that s_{k_j} is the smallest club that satisfies this condition. Now single-crossing property ensures that the median (take the leftmost median if k_1 is even) voter m strictly prefers s_{k_j} to s_{k_1} ; also note that m did not strictly prefer s_{k_r} to s_{k_1} for $r < j$. Now consider state s_{k_j} ; clearly the median voter (or both of them) have number greater than or equal to m ; equality is only possible if $s_{k_j} = s_{k_1} + 1$ and so there are no clubs in between. This means that none of the median voters can strictly prefer s_{k_r} with $r < j$ to s_{k_j} , and therefore such s_{k_r} cannot obtain a majority. Evidently, the same holds for median voters in clubs s_{k_q} with $q > j$. So, we can take clubs s_{k_r} with $r < j$ out of consideration, because if the median voter of club s_{k_q} with $q \geq j$ weakly prefers s_{k_r} , $r < j$, to s_{k_q} , then he strongly prefers s_{k_j} . We can repeat the procedure above; since the number of clubs is finite, we will eventually find a club that is better than any other club in \mathcal{Q} . As for part (b), for the first part it is sufficient to take a median voter in state z and pick his most preferred club among the set of clubs under consideration. For the second part, it suffices to notice that if $s_l \succ_{s_k} s_k$ and $s_j \succ_{s_k} s_l$, for some s_k, s_l , and s_j , then the median voter (or both of them if there are two median voters in state s_k) strictly prefer s_l to s_k and s_j to s_l , and therefore strictly prefer s_j to s_k . But then the single-crossing condition implies that a majority of voters (in s_k) prefer s_j to

s_k , so $s_j \succ_{s_k} s_k$. Therefore, if $s_l \succ_{s_k} s_k$ and $s_j \not\succeq_{s_k} s_k$, then $s_j \not\succeq_{s_k} s_l$, so Assumption 2 holds. It is straightforward to check that Assumption 2(b)* also holds, and one can similarly check that for median voter rule the assumptions are satisfied as well.

(Part 2) The existence of mapping ϕ immediately follows from Theorem 1. Suppose all clubs have odd size; let us show that Assumption 3 holds. Suppose the current state is s_j , and $s_k \succ_{s_j}^{maj} s_j$, $s_l \succ_{s_j}^{maj} s_j$. Then consider median voter $(j + 1) / 2$; because of the single-crossing condition, if he prefers s_k to s_l , then $s_k \succ_{s_j}^{maj} s_l$, and vice versa. Therefore, if we pick among the two clubs s_k and s_l the one he prefers, we show that Assumption 3 is satisfied, and thus mapping ϕ is unique, again by Theorem 1. Note that this proof highlights the reason for possible non-uniqueness of mapping ϕ when even-sized clubs are possible: in that case, there are two median voters, and if there are two clubs that these median voters like best, but disagree over which one is better, then ϕ could map the initial club into any one of these two.

(Part 3) Suppose that β is close to 1 and a pure strategy Markov Voting Equilibrium (MVE) of the Modified Roberts game (which exists by hypothesis) maps each club s to the next-period club $\chi(s)$. Roberts's argument that there are no cycles in the MVE continues to apply. Therefore, with $\chi(s)$ iterated, the sequence $\{\chi^n(s)\}$ will converge in no later than after $N - 1$ iterations. Let $\psi(s) = \chi^{N-1}(s)$ for any club s . We will now use the definition of MVE to show that mapping ψ satisfies Axioms 1–3.

Axiom 3 is trivially satisfied: for any club s , $\psi(\psi(s)) = \chi^{2(N-1)}(s) = \psi(s)$. Take Axiom 1. Suppose it does not hold; given that we assumed no indifferences, this means that for some s , the set of players in club s who prefer $\psi(s)$ to s does not constitute a majority, which, since club size is odd, means that a majority prefers s to $\psi(s)$. But then any player i in this majority has $V_i(\chi(s), \chi(\cdot)) > V_i(s, \chi(\cdot))$, because the continuation value starting from $\chi(s)$ is arbitrarily close to $w_{\psi(s)}(i)$ since β is close to 1, and therefore is worse than $w_s(i)$, making it worthwhile for the majority to stay an additional period in s . This contradicts that χ is a transition rule in a MVE, therefore, 1. The proof that Axiom 2 holds is: if some state z such that $\psi(z) = z$ is preferred to $\psi(s)$ by a majority, then this majority would be better off from switching to state z .

Since Axioms 1–3 are satisfied, mapping ψ must coincide with the unique mapping ϕ from Part 2 that satisfies these Axioms. This completes the proof. ■

Proof of Proposition 2. (Part 1) Assumption 1 is satisfied both for voting rules and constitutions, because in either case $b > N/2$ and we have either a majority or a supermajority voting rule for any state (and such rules automatically satisfy 1). To check that Assumption 2 holds, we use the fact that preferences satisfy single-crossing. That is, if player i prefers voting

rule a' to a , where $a' > a$ (for constitutions: if i prefers constitution (a', b') to (a, b) where $a' > a$) then player $j > i$ also does. Now using exactly the same argument as in the proof of Part 1 of Proposition 1, we obtain that Assumption 2(a) and Assumptions 2(b) and 2(b)* hold in this case as well.

(Part 2) This follows from Part 1 and Theorem 1.

(Part 3) Take a stable voting rule a . Suppose that it is not a stable point of mapping ϕ_v , then $\phi_v(a) = a' \neq a$. Then Axiom 1 implies that $a' \succ_a a$, that is, a winning coalition in a (i.e., at least a voters) prefer a' to a . But this means that a is not a stable voting rule in the sense of Barbera and Jackson, thus yielding a contradiction.

(Part 4) The result that any stable constitution is a stable point of mapping ϕ_c may be proved exactly as for the case of voting rules in Part 3. Now take any non-stable constitution (a, b) ; let us prove that $\phi_c((a, b)) \neq (a, b)$. Consider the set of constitutions $\mathcal{Q} = \{(a', b')\}$ such that $(a', b') \succ_{(a, b)} (a, b)$; since (a, b) is unstable, this set is non-empty. Note that if $(a', b') \in \mathcal{Q}$, then $(a', N) \in \mathcal{Q}$ (because the second part of the pair of rules does not enter the utility directly). Now take some player i and $(a', b') \in \mathcal{Q}$ that is most preferred by i among the states within \mathcal{Q} (or one of such states if there are several of these). Consider state $(a', N) \in \mathcal{Q}$. First, since it lies in \mathcal{Q} , $(a', N) \succ_{(a, b)} (a, b)$. Second, this state is ϕ_c -stable: indeed, if it were not the case, we would have some other $(a'', b'') \succ_{(a', N)} (a', N)$. This means that each player prefers (a'', b'') to (a', N) , which of course implies that at least a players prefer (a'', b'') to (a, b) , so $(a'', b'') \in \mathcal{Q}$. But there is player i who at least weakly prefers (a', b') (and therefore (a', N) , which is the same as far as immediate payoffs are concerned) to any other element in \mathcal{Q} . This means that such (a'', b'') does not exist, and state (a', N) is stable. Axiom 2 then implies that $\phi_c(a, b)$ cannot equal (a, b) , since state (a', N) is ϕ_c -stable and is preferred to (a, b) . This completes the proof. ■

Proof of Proposition 3. (Part 1) Take any coalition X . If some subcoalition Y which has a weighted α -majority, then any coalition Z such that $Y \subset Z \subset X$ also does. Moreover, since $\alpha \geq 1/2$, any such coalitions must intersect. Hence, Assumption 1' is satisfied. Assumption 2'(a) is satisfied trivially: if \mathcal{Q} is such that any transitions between its elements are allowed, then it is a singleton. To verify Assumption 2'(b), within the set of coalitions \mathcal{Q} we can simply pick the coalition with the least total power. Assumption 3' also holds because among the two coalitions preferred to the current one, the weighted α -majority would always prefer the one with the least total power. Finally, if transitions $X \rightsquigarrow Y$ and $Y \rightsquigarrow Z$ are feasible, then $Y \subset X$ and $Z \subset Y$, hence $Z \subset X$, and therefore transition $X \rightsquigarrow Z$ is feasible, too.

(Part 2) The existence and uniqueness result follows from Part 1 and Theorem 3. To show the equivalence result, one needs to check that if the outcome mapping ϕ_{elim} satisfies Axioms 1'–3', it also satisfies Axioms 1–4 in Acemoglu, Egorov, and Sonin (2008), which, as we proved there, is unique, and in the generic case we consider, single-valued. Then any initial state s_0 is mapped by the mapping ϕ_{elim} to the same state as by the mapping from that paper.

(Part 3) Assumption 1 is not changed as compared to Part 1, so there is nothing to prove. To prove that Assumption 2(a) holds, for any set of states \mathcal{Q} take the one with the least total power. The proofs for Assumption 2(b) and 3 are unchanged from ones for Part 1. Consequently, by Theorem 1, there exists a unique mapping ϕ that satisfies Axioms 1–3. To show that ϕ may be different from ϕ_{elim} , consider the following example. There are 4 players A, B, C, D with $\gamma_A = 3$, $\gamma_B = 4$, $\gamma_C = 5$, and $\gamma_D = 4.5$, and consider the case of weighted majority voting $\alpha = 1/2$. Both ϕ and ϕ_{elim} map any one-player coalition to itself, and any two-person coalition to the one which includes the stronger player only. Consider coalition $\{A, B, C\}$. It is easy to see that $\phi_{\text{elim}}(\{A, B, C\}) = \{A, B, C\}$, but $\phi(\{A, B, C\}) = \{A, B, D\}$. This shows that $\{A, B, C\}$ is stable under ϕ_{elim} but is unstable under ϕ . This completes the proof. ■

Appendix B

Example 6 There are 3 players, $\mathcal{I} = \{1, 2, 3\}$, and 3 states, $\mathcal{S} = \{A, B, C\}$. Players' preferences satisfy $w_A(1) > w_B(1) > w_C(1)$, $w_B(2) > w_C(2) > w_A(2)$, and $w_C(3) > w_A(3) > w_B(3)$ (for example, $w_A(1, 2, 3) = (10, 5, 8)$, $w_B(1, 2, 3) = (8, 10, 5)$, $w_C(1, 2, 3) = (5, 8, 10)$). Winning coalitions are given by $\mathcal{W}_A = \{C \in \mathcal{C} : 3 \in C\}$, $\mathcal{W}_B = \{C \in \mathcal{C} : 1 \in C\}$, $\mathcal{W}_C = \{C \in \mathcal{C} : 2 \in C\}$ (in other words, states A, B, C have dictators 1, 2, 3, respectively). We then have $A \succ_B B$, $B \succ_C C$, $C \succ_A A$, so Assumption 2(a) is violated.

It is easy to see that there is no dynamically stable state in the dynamic game in this case. To see this, suppose that state A is dynamically stable, then state B is not, since player 1 would enforce transition to A . Therefore, state C is stable: player 2, who is the dictator in C , knows that a transition to B will lead to A , which is worse than C . However, then player 3, knowing that C is stable, will have an incentive to move from A to C . In equilibrium this deviation should not be profitable, but it is; hence, there is no equilibrium where A is stable. Now, given the transition costs, there is no MPE in pure strategies, since if no state is dynamically stable, the players would benefit from blocking any single transition in any single state.

Let us now formally show that there is no mapping ϕ that satisfies Axioms 1–3. Assume that there is such mapping ϕ . By Axiom 3, there is a stable state (for any state s , $\phi(s)$ is stable). Without loss of generality, suppose that A is such a state: $\phi(A) = A$. Then state C is not stable: if it were, we would obtain a contradiction with Axiom 2, since $C \succ_A A$. If C is not stable, then either $\phi(C) = A$ or $\phi(C) = B$. The first is impossible by Axiom 1, since player 2, who is a member of any winning coalition in C , has $w_C(2) > w_A(2)$. Therefore, $\phi(C) = B$, and by Axiom 3, $\phi(B) = B$. But we have $A \succ_B B$ and $\phi(A) = A$; this means, by Axiom 2, that $\phi(B) = B$ cannot hold. This contradiction shows that with these preferences, there is no mapping ϕ that satisfies Axioms 1–3.

Example 7 There are 3 players, $\mathcal{I} = \{1, 2, 3\}$, and 4 states, $\mathcal{S} = \{A, B, C, D\}$. Players' preferences satisfy $w_A(1) > w_B(1) > w_C(1) > w_D(1)$, $w_B(2) > w_C(2) > w_A(2) > w_D(2)$, and $w_C(3) > w_A(3) > w_B(3) > w_D(3)$ (for example, $w_A(1, 2, 3) = (10, 5, 8)$, $w_B(1, 2, 3) = (8, 10, 5)$, $w_C(1, 2, 3) = (5, 8, 10)$, $w_D(1, 2, 3) = (4, 4, 4)$). Winning coalitions are given by $\mathcal{W}_A = \mathcal{W}_B = \mathcal{W}_C = \{\mathcal{I}\} = \{\{1, 2, 3\}\}$, $\mathcal{W}_D = \{\{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$ (in other words, in states A, B, C there is unanimity, while in state D there is majority voting rule). We then have $A \succeq_D B$, $B \succeq_D C$, $C \succeq_D A$ with at least one (actually, all) relations supported by players for whom neither A nor B nor C is the worst possible outcome, and $A \approx B \approx C \approx A$, so

Assumption 2(b) is violated. Assume, in addition, that $K_D = 3$, and $\pi_D(1) = C$, $\pi_D(2) = B$, $\pi_D(3) = A$.

In this case, states A, B, C are dynamically stable: evidently, player who receives 10 (1, 2, 3, respectively) will block transition to either of the states. Consider state D ; it is easy to see that it is not dynamically stable. Indeed, if it were, then all three players would be better off from transition to either of the three other states A, B, C , so they must vote for any such proposal in equilibrium. Now that it is not dynamically stable, we must have that some of proposals C, B, A are accepted in equilibrium. Suppose that A is accepted, then B may not be accepted (because two players, 1 and 3, strictly prefer A to B), and therefore C must be accepted (because two players, 2 and 3, strictly prefer C to A). But then A may not be accepted, as players 2 and 3 would prefer to have it rejected so that C is accepted in the next period, and by Lemma 1(c) A must be rejected in the equilibrium. This contradicts our assertion that A is accepted, and we would obtain a similar contradiction if we assumed that some other proposal is accepted. Hence, there is no MPE in pure strategies in this case.

Let us now show that there is no mapping ϕ that satisfies Axioms 1–3. Assume that there is such mapping ϕ . Since for each of the states A, B, C there is no state that is preferred to it by all three players, then Axiom 1 implies that $\phi(A) = A$, $\phi(B) = B$, and $\phi(C) = C$. Consider state D . If $\phi(D) = D$, this would violate Axiom 2, since, for instance, state A satisfies $A \succ_D D$ and $\phi(A) = A$. Hence, $\phi(D) \neq D$; without loss of generality assume $\phi(D) = A$. But then state C satisfies $C \succ_D A$, $C \succ_D D$, and $\phi(C) = C$. By Axiom 2 we cannot have $\phi(D) = A$, which brings us to a contradiction. We have thus proved that there does not exist mapping ϕ that satisfies Axioms 1–3.

Example 8 There are 2 players, $I = \{1, 2\}$, and 3 states, $S = \{A, B, C\}$. Players' preferences satisfy $w_A(1) > w_B(1) > w_C(1)$, $w_B(2) > w_A(2) > w_C(2)$ (for example, $w_A(1, 2) = (5, 3)$, $w_B(1, 2) = (3, 5)$, $w_C(1, 2) = (1, 1)$). Winning coalitions are given by $\mathcal{W}_A = \mathcal{W}_B = \mathcal{W}_C = \{\mathcal{I}\} = \{\{1, 2\}\}$ (in other words, there is a unanimity voting rule in all states A, B, C). It is then easy to see that Assumptions 1 and 2(a,b) are satisfied, while Assumption 3 is violated (both A and B are preferred to C , but neither $A \succ_C B$ nor $B \succ_C A$).

It is then easy to see that in this case, there exist two mappings, ϕ_1 and ϕ_2 , which satisfy Axioms 1–3. Let $\phi_1(A) = \phi_1(C) = A$ and $\phi_1(B) = B$. Let $\phi_2(A) = A$ and $\phi_2(B) = \phi_2(C) = B$. Mappings ϕ_1 and ϕ_2 differ in only that the first maps state C to state A , and the second maps state C to state B . It is straightforward to verify that ϕ_1 and ϕ_2 satisfy Axioms 1–3, and also that no other mapping satisfies these Axioms. Evidently, the set of stable states under

these two mappings satisfy $\mathcal{D}_{\phi_1} = \{A, B\} = \mathcal{D}_{\phi_2}$ (so stable states of mapping ϕ_1 and mapping ϕ_2 are the same). This is not a coincidence, but rather a general result proved in Theorem 1.

Example 9 There are 2 players, $I = \{1, 2\}$, and 3 states, $S = \{A, B, C\}$. Players' preferences satisfy $w_C(1) > w_B(1) > w_A(1)$, $w_B(2) > w_A(2) > w_C(2)$ (for example, $w_A(1, 2) = (3, 3)$, $w_B(1, 2) = (6, 4)$, $w_C(1, 2) = (9, 1)$). Winning coalitions are given by $\mathcal{W}_A = \{\mathcal{I}\} = \{\{1, 2\}\}$, $\mathcal{W}_B = \mathcal{W}_C = \{\{1\}, \{1, 2\}\}$ (in other words, there is a unanimity voting rule in state A , while in states B and C player 1 is the dictator). Clearly, state A is Pareto superior to state B (and better for winning coalition $\{A, B\}$).

However, the unique mapping ϕ that satisfies Axioms 1–3 has $\phi(A) = A$, $\phi(B) = \phi(C) = C$. To see this, note first that there is no state which is better than C to player 1. Therefore, by Axiom 1, we must have $\phi(C) = C$. Now take state B . Axiom 1 implies that either $\phi(B) = B$ or $\phi(B) = C$, but $\phi(B) = B$ is ruled out by Axiom 2, so $\phi(B) = C$ and state B is unstable. Now, by Axiom 1, either $\phi(A) = A$ or $\phi(A) = B$, but at the same time, by Axiom 3, either $\phi(A) = A$ or $\phi(A) = C$. Both these conditions are satisfied only if $\phi(A) = A$, so A is a stable state.

We see here that if the game starts with state A as the initial state, then it will remain there (A is a dynamically stable state). This is rather surprising, since state B is preferred to A by all players. However, this is perfectly consistent with the spirit and the results of this paper. Indeed, state B is unstable, as everyone knows that once it becomes the current state, there will be a further transition to state C . Consequently, when players compare whether or not to move from state A to state B , they effectively compare their benefits from states A and C , not A and B . With this in mind, the transition still seems profitable for player 1 (he prefers C to A). However, player 2 is worse off from such transition, as he would receive a lower payoff in state C than in state A). As a result, no transition from state A to any other state may happen in an equilibrium, because player 2 has a veto power in state A which he will exercise. Note that if player 2 had veto power in state B as well, he would be able to block the transition from B to C , and in that case, state A would no longer be stable.

References

- Acemoglu, Daron (2003) "Why Not a Political Coase Theorem? Social Conflict, Commitment, and Politics," *Journal of Comparative Economics*, 31, 620–652.
- Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin (2008) "Coalition Formation in Non-democracies," *Review of Economic Studies*, forthcoming.
- Acemoglu, Daron and James Robinson (2000) "Why Did The West Extend The Franchise? Democracy, Inequality, And Growth In Historical Perspective," *Quarterly Journal of Economics*, 115(4), 1167-1199.
- Acemoglu, Daron and James Robinson (2006a) *Economic Origins of Dictatorship and Democracy*, Cambridge University Press.
- Acemoglu, Daron and James Robinson (2006b) "Economic Backwardness in Political Perspective" *American Political Science Review*.
- Alesina, Alberto, Ignazio Angeloni, and Federico Etro (2005) "International Unions," *American Economic Review*, vol. 95(3), pages 602-615.
- Austen-Smith, David and Jeffrey Banks (1988) "Elections, Correlations and Legislative Outcomes," *American Political Science Review* 82: 405-422.
- Barbera, Salvador, Michael Maschler, and Jonathan Shalev (2001) "Voting for Voters: A Model of the Electoral Evolution," *Games and Economic Behavior*, 37: 40-78.
- Barbera, Salvador, and Matthew Jackson (2004) "Choosing How to Choose: Self-Stable Majority Rules and Constitutions," *Quarterly Journal of Economics*, 119(3), 1011-1048.
- Baron, David (1991) "A Spatial Bargaining Theory of Government Coalition Formation in Parliamentary Systems", *American Political Science Review*, 83: 993-967.
- Baron, David and John Ferejohn (1989) "Bargaining in Legislatures," *American Political Science Review* 83: 1181-1206.
- Bordignon, Massimo and Sandro Brusco (2006) "On Enhanced Cooperation," mimeo.
- Bourguignon, François and Thierry Verdier (2000) "Oligarchy, Democracy, Inequality and Growth," *Journal of Development Economics*, 62, 285-313.
- Buchanan, James (1965) "An Economic Theory of Clubs," *Economica*, 32, 125, pp. 1-14.
- Burkart, Mike and Klaus Wallner (2000) "Club Enlargement: Early Versus Late Admittance," mimeo.
- Chwe, Michael S. Y. (1994) "Farsighted Coalitional Stability," *Journal of Economic Theory*, 63: 299-325.
- Fearon, James (1996) "Bargaining Over Objects that Influence Future Bargaining Power,"

mimeo.

Fearon, James (1998) "Commitment Problems and the Spread of Ethnic Conflict", in David Lake and Donald Rothchild, eds., *The International Spread of Ethnic Conflict: Fear, Diffusion, and Escalation*, Princeton: Princeton Univ. Press.

Fearon, James (2004) "Why Do Some Civil Wars Last so Much Longer Than Others," *Journal of Peace Research*, 41, 275-301.

Giovannoni, Francesco and Toke Aidt (2004) "Constitutional Rules," mimeo.

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

Jack, William and Roger Lagunoff (2006) "Dynamic Enfranchisement," *Journal of Public Economics*, vol. 90(4-5), pages 551-572.

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

Jehiel, Philippe and Suzanne Scotchmer (2001), "Constitutional Rules of Exclusion in Jurisdiction Formation," *Review of Economic Studies*, 68:393-413.

Lizzeri, Alessandro and Nicola Persico (2004) "Why Did the Elites Extend the Suffrage? Democracy and the Scope of Government, With an Application to Britain's 'Age of Reform'," *Quarterly Journal of Economics*, vol. 119(2), pages 705-763.

Messner, Matthias and Matthias Polborn (2004) "Voting on Majority Rules," *Review of Economic Studies*, 71(1), 115-132

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

Norman, Peter (2002) "Legislative Bargaining and Coalition Formation," *Journal of Economic Theory* 102: 322-353.

Powell, Robert (1998) *In the Shadow of Power: States and Strategies in International Politics*, Princeton Univ. Press, Princeton, NJ.

Rajan, Raghuram and Luigi Zingales (2000) The Tyranny of Inefficient: An Enquiry into the Adverse Consequences of Power Struggles, *Journal of Public Economics*, Vol. 76, no.3, 521-558.

Ray, Debraj (2008) *A Game-Theoretic Perspective on Coalition Formation*, Oxford University Press.

Riley, John (1979) "Informational Equilibrium," *Econometrica*, 47, 331-59.

Roberts, Kevin (1999) "Dynamic Voting in Clubs," mimeo.

Robinson James A. (1997) "When Is the State Predatory?" mimeo.

Rothschild, Michael and Joseph Stiglitz (1976) "Equilibrium in Competitive Insurance Markets:

An Essay in the Economics of Imperfect Information,” *Quarterly Journal of Economics*, 80, 629-49.

Schwarz, Michael and Konstantin Sonin (2008) “A Theory of Brinkmanship, Conflicts, and Commitment”, *Journal of Law, Economics, and Organization*, forthcoming.

Scotchmer, Suzanne (2002) “Local Public Goods and Clubs,” In Alan Auerbach and Martin Feldstein (editors) *Handbook of Public Economics*, IV. Chapter 29, Amsterdam: North-Holland Press, 1997-2042.