NBER WORKING PAPER SERIES

DYNAMICS AND STABILITY OF CONSTITUTIONS, COALITIONS, AND CLUBS

Daron Acemoglu
Georgy Egorov
Konstantin Sonin

Working Paper 14239
http://www.nber.org/papers/w14239

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2008

Daron Acemoglu gratefully acknowledges financial support from the National Science Foundation and the Canadian Institute for Advanced Research. We thank John Duggan, Matthew Jackson, Robert Powell, and conference and seminar participants at the ASSA 2008 (New Orleans), the 2008 Game Theory World Congress, Berkeley, California Institute of Technology, Harvard and Princeton for useful comments. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Dynamics and Stability of Constitutions, Coalitions, and Clubs
Daron Acemoglu, Georgy Egorov, and Konstantin Sonin
NBER Working Paper No. 14239
August 2008
JEL No. C71,D71,D74

ABSTRACT

A central feature of dynamic collective decision-making is that the rules that govern the procedures for future decision-making and the distribution of political power across players are determined by current decisions. 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 dynamic problems. Under relatively natural acyclicity assumptions, we provide a complete characterization of dynamically stable states as functions of the initial state and determine conditions for their uniqueness. We show how this framework can be applied in political economy, coalition formation, and the analysis of the dynamics of clubs. The explicit characterization we provide highlights two intuitive features of dynamic collective decision-making: (1) a social arrangement is made stable by the instability of alternative arrangements that are preferred by sufficiently many members of the society; (2) efficiency-enhancing changes are often resisted because of further social changes that they will engender.

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

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

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

1 Introduction

Consider the problem of a society choosing its constitution. Naturally, the current rewards from adopting a specific constitution will influence 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 political control. 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, coalition formation, organizational economics, and industrial organization have a structure resembling this example of constitutional choice.

We develop a general framework for the analysis of dynamic group-decision-making over constitutions, coalitions, and clubs. 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 and political decisions are made. It determines stage payoffs and also how the society can determine its future states (constitutions), for example, which subsets of individuals can change the constitution. Our focus is on (Markov perfect) equilibria of this dynamic game when individuals are sufficiently forward-looking. Under natural acyclicity assumptions, which rule out Condorcet-type cycles, we prove the existence and characterize the structure of (*dynamically*) *stable states*, which are defined as states that arise and persist. An equilibrium is represented by a mapping ϕ, which designates the dynamically stable state $\phi(s_0)$ as a function of the initial state s_0. We show that the set of dynamically stable states is largely independent of the details of agenda-setting and voting protocols.

Although our main focus is the analysis of the noncooperative game outlined in the previous paragraph, we first start with an axiomatic characterization of stable states. This characterization relies on the observation that sufficiently forward-looking individuals will not wish to support change towards a state (constitution) that might ultimately lead to another, less preferred state. This observation is encapsulated in a simple *stability axiom*. We also introduce two other natural axioms ensuring that individuals do not support changes that will give them lower utility. We characterize the set of mappings, Φ, that are consistent with these three axioms and provide conditions under which there exists a unique $\phi \in \Phi$ (Theorem 1). Even when Φ is not a singleton, the sets of stable states according to any two $\phi_1, \phi_2 \in \Phi$ are identical.

Our main results are given in Theorem 2. This theorem shows that for any agenda-setting and voting protocol the equilibria of the dynamic game we outline can be represented by some $\phi \in \Phi$ and that for any $\phi \in \Phi$, there exists a protocol such that the equilibrium will be represented by ϕ.[1] This means that starting with initial state s_0, any equilibrium leads to a dynamically stable state $\phi(s_0)$ for some $\phi \in \Phi$. Naturally, when such ϕ is unique, all equilibria result in the uniquely-defined dynamically stable $\phi(s_0)$.

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.

At the center of our approach is the natural lack of *commitment* in dynamic decision-making problems—those that gain additional decision-making power as a result of a reform cannot commit to refraining from further choices that would hurt the initial set of decision-makers. This lack of commitment, together with forward-looking behavior, is at the root of the general characterization result provided in these theorems. It also leads to two simple intuitive results:

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 constituency. To understand why certain social arrangements are stable, we must thus study the instabilities that changes away from these arrangements would unleash.[2]

2. Dynamically stable states can be inefficient—in the sense that they may be Pareto dominated by the payoffs in another state.

Our final general result, Theorem 3, provides sufficient conditions for the acyclicity assumptions in Theorems 1 and 2 to hold when states belong to an ordered set (e.g., when they are a subset of $\mathbb{R}$). In particular, it shows that these results apply when (static) preferences satisfy a single-crossing property or are single peaked (and some minimal assumptions on the structure of winning coalitions are satisfied). This theorem makes our main results easy to apply in a variety of environments. We also show that Theorems 1 and 2 apply in a range of situations in which states do not belong to an ordered set.

The next two examples provide simple illustrations of the dynamic trade-offs emphasized by our approach.

[1] An additional assumption in the analysis of this game 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 (see also Examples 3 and 4 in Appendix C).

[2] This result also shows that, in contrast to Riker's (1962) emphasis in the context of political coalition formation games, the equilibrium will typically not involve a "minimum winning coalition," because the state corresponding to this coalition is generally unstable.

Example 1 (Inefficient Inertia) Consider a society consisting of two individuals, E and M. E represents the elite and initially holds power, and M corresponds to the middle class. There are three states: (1) absolutist monarchy a, in which E rules, with no checks and no political rights for M; (2) constitutional monarchy c, in which M has greater security and is willing to invest; (3) democracy d, where M becomes more influential and the privileges of E disappear. Suppose that stage payoffs satisfy

$$w_E(d) < w_E(a) < w_E(c), \text{ and } w_M(a) < w_M(c) < w_M(d).$$

In particular, $w_E(a) < w_E(c)$ 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. To simplify the discussion, suppose that starting in both regimes c and d, M decides next period's regime. In terms of the notation introduced above, this implies that d is a dynamically stable state, and $\phi(d) = d$. In contrast, c is not a dynamically stable state, since starting from c, there will be a transition to d and thus $\phi(c) = d$. Therefore, if, starting in regime a, E chooses a reform towards c, this will lead to d in the following period, and thus give E a discounted payoff of

$$U_E(\text{reform}) = w_E(c) + \beta \frac{w_E(d)}{1-\beta}.$$

In contrast, if E decides to stay in a forever, its payoff is $U_E(\text{no reform}) = w_E(a)/(1-\beta)$. If β is sufficiently small, then $U_E(\text{no reform}) < U_E(\text{reform})$, and reform will take place. However, when players are forward looking and β is large, then $U_E(\text{no reform}) > U_E(\text{reform})$. In this case, the unique dynamically stable state starting with a is a—that is, $\phi(a) = a$.

This example, when players are sufficiently forward-looking, illustrates both of the intuitive results mentioned above. First, state a is made stable by the instability of another state, c, that is preferred by those who are powerful in a. Second, both E and M would be strictly better off in c than in a, so the stable state starting from a is Pareto inefficient. It also illustrates another general insight: the set of stable states is larger when players are forward-looking (when β is small, only d is stable, whereas when β is large, both a and d are stable).

A similar game can be used to model the implications of concessions in wars. For example, a concession that increases the payoffs to both warring parties may not take place because it will change the future balance of power. It 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 the CEO would then become powerful and reduce their privileges.[3]

Example 2 (Voting in Clubs) Consider the problem of voting in clubs. The society consisting of N individuals. A club is a subset of the society. Each individual i receives a stage payoff $w_i(s_t)$, which is as a function of the current club s_t, and current club members decide (according to some voting rule) tomorrow's club s_{t+1}. The seminal unpublished paper by Roberts (1999) studies a special case of this environment, where individuals are ordered, $i = 1, 2, ..., N$, any club s_t must take the form $x_k = \{1, ..., k\}$ for some $k = 1, 2, .., N$, and decisions are made by majoritarian voting. Under a range of additional assumptions Roberts establishes the existence of mixed-strategy (Markovian) equilibria and characterizes some of their properties. Our model nests a more general version of this environment and enables us to establish the existence of a unique dynamically stable club (and a pure-strategy equilibrium) under weaker conditions. In addition, our approach allows a complete characterization of dynamically stable states and clarifies the reasons for potential Pareto inefficiency.

These examples illustrate some of the possible applications of our approach. We view the rich set of environments that are covered by our model and the relative simplicity of the resulting dynamic stable states as its major advantages. Both our specific results 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. Barbera, Maschler, and Shalev (2001) study a dynamic game of club formation in which any member of the club can unilaterally admit a new agent. The recent ambitious paper by Lagunoff (2006), which constructs a general model of political reform and relates reform to the time-inconsistency of induced social rules, is another precursor. Acemoglu and Robinson's (2000, 2006a) and Lizzeri and Persico's (2004) analyses of franchise extension and Barbera and Jackson's (2004) model of constitutional stability are on related themes as well. How these papers can be viewed as applications of our general framework is discussed in Section 6.

The two papers most closely related to our work are Chwe (1994) and Gomes and Jehiel

[3] Ideas related to this example have been discussed in a number of different 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 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.

(2005). Chwe studies a model where payoffs are determined by states and there are exogenous rules governing transitions from one state to another. Chwe demonstrates the relationship between two distinct notions from cooperative game theory, the consistent and stable sets. However, in Chwe's setup, neither a noncooperative analysis nor characterization results are possible, while such results are at the heart of our paper. The link between Chwe's consistent sets and our dynamically stable states is discussed further below. Gomes and Jehiel study a related environment with side payments. They 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 make transitions to non-stable states in our paper. They also show that equilibrium may be inefficient when the discount factor is small. In contrast, in our game Pareto dominated outcomes are not only possible in general, but they may emerge as unique equilibria and are more likely when discount factors are large (as illustrated by Example 1). 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 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, and 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. Alesina, Angeloni, and Etro (2005) apply a simplified version of Roberts's model to the enlargement of the EU and Bordignon and Brusco (2003) study the role of "enhanced cooperation agreements" in the dynamics of EU enlargement.

The rest of the paper is organized as follows. Section 2 introduces the general environment. Section 3 motivates and presents our axiomatic analysis, which also acts as a preparation for our noncooperative analysis. In Section 4, we prove the existence of a (pure-strategy) Markov perfect equilibrium of the dynamic game for any agenda setting and voting protocol and establish the equivalence between these equilibria and the axiomatic characterization in Section 3. Section 5 shows how the results of Sections 3–4 can be applied when states belong to an ordered set. Section 6 discusses a range of applications of our framework, including the two examples presented above. Section 7 concludes. Appendix A presents the main proofs omitted from the text. Appendix B and C, which are not for publication, contain a number of generalizations, additional results, examples, and some omitted proofs.

2 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 $\mathcal{S}$. Throughout the paper, $|X|$ denotes the number of elements of set X, so $|\mathcal{I}|$ and $|\mathcal{S}|$ denote the number of individuals and states, respectively. States represent both different institutions affecting players' payoffs and procedures for decision-making (e.g., the identity of the ruling coalition in power, the degree of supermajority, or 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 different 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 determined endogenously. A nonempty 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 $\left(\{w_i(s)\}_{i\in\mathcal{I}}, \mathcal{W}_s\right)$. Here, for each state $s \in \mathcal{S}$, $w_i(s)$ is a (strictly) positive stage payoff assigned to each individual $i \in \mathcal{I}$. The restriction that $w_i(s) > 0$ is a normalization, making 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. We use $\mathcal{W}_s$ to model political institutions in state s. This allows us to summarize different political procedures, such as weighted majority or supermajority rules, in an economical fashion. For example, if in state s a majority is required for decision-making, $\mathcal{W}_s$ includes all subsets of $\mathcal{I}$ that form a majority; if in state s individual i is a dictator, $\mathcal{W}_s$ includes all coalitions that include i.[4] Since $\mathcal{W}_s$ is a function of the state, the procedure for decision-making can vary across states.

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:*

(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 $X \cap Y \neq \varnothing$.

Part (a) simply states that if some coalition X is winning in state s, then increasing the size of the coalition will not reverse this. This is a natural assumption for almost any decision rule.

[4] Notice that $\mathcal{W}_s$ or the political rules do not specify certain institutional details, such as who makes proposals, how voting takes place and so on. These are specified by the agenda-setting and voting protocols of our dynamic game. We will show that these only have a limited effect on equilibrium outcomes, justifying our focus on $\mathcal{W}_s$ as a representation of "political rules".

Part (b) rules out the possibility that two disjoint coalitions are winning in the same state, thus imposing a form of (possibly weighted) majority or supermajority rule. If $\mathcal{W}_s = \varnothing$, then state s is *exogenously stable*. None of our existence or characterization results depend on whether there exist exogenously stable states.

The following binary relations on the set of states $\mathcal{S}$ will be useful for the rest of our analysis. For $x, y \in \mathcal{S}$, we write

$$x \sim y \iff \forall i \in \mathcal{I} : w_i(x) = w_i(y). \tag{1}$$

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

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

Intuitively, $y \succeq_z x$ means that there exists a coalition of players that is winning (in z) with each of its members weakly preferring y to x. Note three important features about $\succeq_z$. First, it only contains information about stage payoffs. In particular, $w_i(y) \geq w_i(x)$ does *not* mean that individual i will prefer a switch to state y rather than x. Whether or not he does so depends on the continuation payoffs following such a switch. Second, the relation $\succeq_z$ does not presume any type of coordination or collective decision-making among the members of the coalition in question. It simply records the existence of such a coalition. Third, the relation $\succeq_z$ is conditioned on z since whether the coalition of players weakly preferring y to x is winning depends on the set of winning coalitions, which is state dependent. With a slight abuse of terminology, if (2) holds, we say that y is *weakly preferred* to x in z. In light of the preceding comments, this neither means that all individuals prefer y to x nor that there will be a change from state x to y in the dynamic game—it simply designates that there exists a winning coalition of players, each obtaining a greater stage payoffs in y than in x. Relation $\succ_z$ is defined similarly by

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

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

Relation $\sim$ clearly defines equivalence classes; if $x \sim y$ and $y \sim z$, then $x \sim z$. In contrast, the binary relations $\succeq_z$ and $\succ_z$ need not even be transitive. Nevertheless, for any $x, z \in \mathcal{S}$, we have $x \not\succ_z x$, and whenever $\mathcal{W}_z$ is nonempty, we also have $x \succeq_z x$. Finally, from Assumption 1 we have that 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 $x \not\succ_z y$.

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

Assumption 2 ***(Payoffs)*** *Payoffs* $\{w_i(s)\}_{i \in \mathcal{I}, s \in \mathcal{S}}$ *satisfy the following properties:*

(a) For any sequence of states $s_1, s_2, \ldots, s_k$ in $\mathcal{S}$,

$$s_{j+1} \succ_{s_j} s_j \text{ for all } 1 \leq j \leq k-1 \Longrightarrow s_1 \not\succ_{s_k} s_k.$$

(b) For any sequence of states $s, s_1, \ldots, s_k$ in $\mathcal{S}$ with $s_j \succ_s s$ for each $1 \leq j \leq k$,

$$s_{j+1} \succ_s s_j \text{ for all } 1 \leq j \leq k-1 \Longrightarrow s_1 \not\succ_s s_k.$$

Assumption 2 plays a major role in our analysis and ensures "acyclicity" (but is considerably weaker than "transitivity"). Part (a) of Assumption 2 rules out cycles of the form $y \succ_x x$, $z \succ_y y$, $x \succ_z z$—that is, a cycle of states (x, y, z) such that in each, a winning coalition of players strictly prefer the next state. Part (b) of Assumption 2 rules out cycles of the form $y \succ_s x$, $z \succ_s y$, $x \succ_s z$.[5] Assumptions 1 and 2 are natural given our focus. Throughout the paper we suppose that they hold. In addition, we sometimes impose the following (stronger) requirement.

Assumption 3 ***(Comparability)*** *For $x, y, s \in \mathcal{S}$ such that $x \succ_s s$, $y \succ_s s$, and $x \not\sim y$, either $y \succ_s x$ or $x \succ_s y$.*

Assumption 3 means that if two states x and y are weakly preferred to s (in s), then y and z are $\succ_s$-comparable. It turns out to be sufficient to guarantee uniqueness of equilibria.[6] This assumption is not necessary for the majority of our results, including the general characterization. Our main results are stated without this assumption and are then strengthened by imposing it.

In each period, each individual maximizes his discounted expected utility:

$$U_i(t) = (1-\beta) \sum\nolimits_{\tau=t}^{\infty} \beta^{\tau} u_i(\tau), \tag{4}$$

where $\beta \in (0, 1)$ is a common discount factor and we can think of $u_i(t)$ as given by the payoff function $w_i(s)$ introduced in Assumption 2 (see, in particular, equation (9) in Section 4). We consider situations in which β is greater than some threshold $\beta_0 \in (0, 1)$ (this threshold is derived as an explicit function of payoffs in Appendix A). We will then characterize the Markov perfect equilibrium (MPE) of this dynamic game and investigate the existence and structure of *dynamically stable states*. As defined more formally in Definition 2, a state s is dynamically stable if there exists a MPE and a finite time T such that this equilibrium involves $s_t = s$ for all $t \geq T$—that is, a dynamically stable state persists in equilibrium.

[5] Neither part of Assumption 2 is implied by the other. Examples 5 and 6 in Appendix C illustrate the types of cycles that can arise when either 2(a) or 2(b) fails.

[6] It is also "necessary" in the sense that if this assumption is dispensed with, it is easy to construct examples with multiple equilibria. Example 7 in Appendix C illustrates this.

3 Axiomatic Characterization

Before specifying the details of agenda-setting and voting protocols, we provide a more abstract (axiomatic) characterization of stable states. This axiomatic analysis has two purposes. First, it illustrates that the key economic forces that arise in the context of dynamic collective decision-making are largely independent of the details of the agenda-setting and voting protocols. Second, the results in this section are a preparation for the characterization of the equilibrium of the dynamic game introduced in the previous section; in particular, our main result, Theorem 2, will make use of this axiomatic characterization.

The key economic insight enabling an axiomatic characterization is that *with sufficiently forward-looking behavior, an individual should not wish to transit to a state that will ultimately lead to another state that gives her lower utility.* This basic insight enables a tight characterization of (*axiomatically*) *stable states.* Theorem 2 in the next section shows the equivalence between the notions of axiomatically and dynamically stable states.

More formally, our axiomatic characterization determines a set of mappings Φ such that for any $\phi \in \Phi$, $\phi : \mathcal{S} \to \mathcal{S}$ assigns an axiomatically stable state $s^{\infty} \in \mathcal{S}$ to each initial state $s_0 \in \mathcal{S}$. We impose the following three axioms on ϕ.

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

Axiom 3 ***(Rationality)*** *If $x, y, z \in \mathcal{S}$ are such that $z \succ_x x$, $z = \phi(z)$, and $z \succ_x y$, then $y \neq \phi(x)$.*

All three axioms are natural in light of what we have discussed above. Axiom 1 requires that the society should not (permanently) move from state x to another state y unless there is a winning coalition that supports this transition. Axiom 2 encapsulates the stability notion discussed above; if some state is not dynamically stable, it cannot be the (ultimate) stable state for any initial state, because there will eventually be a transition away from this state (and thus if mapping ϕ picks state y starting from state x, then it should also pick y starting from y). Axiom 3 imposes the reasonable requirement that if there exists a stable state z preferred to both x and y by winning coalitions in state x, then ϕ should not pick y in x.

All three axioms refer to properties of ϕ, but they are closely related to underlying individual preferences. 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).

We next define the set Φ formally and state the relationship between axiomatically stable states and the mapping ϕ.

Definition 1 ***(Axiomatically Stable States)*** *Let $\Phi \equiv \{\phi : \mathcal{S} \to \mathcal{S}\text{: } \phi \text{ satisfies Axioms } 1\text{–}3\}$. A state $s \in \mathcal{S}$ is* ***(axiomatically) stable*** *if $\phi(s) = s$ for some $\phi \in \Phi$. The set of stable states (fixed points) for mapping $\phi \in \Phi$ is $\mathcal{D}_\phi = \{s \in \mathcal{S}\text{: } \phi(s) = s \text{ for } \phi \in \Phi\}$ and the set of all stable states is $\mathcal{D} = \{s \in \mathcal{S}\text{: } \phi(s) = s \text{ for some } \phi \in \Phi\}$.*

The next theorem establishes the existence of stable states and provides a recursive characterization of such states. It also paves the way for Theorem 2, which shows the equivalence between the equilibrium of the dynamic game in the previous section and the mappings $\phi \in \Phi$.

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

1. *The set Φ is non-empty. That is, there exists a mapping ϕ satisfying Axioms 1–3.*

2. *Any $\phi \in \Phi$ can be recursively constructed as follows. Order the states as $\left\{\mu_1, ..., \mu_{|\mathcal{S}|}\right\}$ such that for any $1 \le j < l \le |\mathcal{S}|$, $\mu_l \not\succ_{\mu_j} \mu_j$ (this is feasible given Assumption 2(a)). Let $\phi(\mu_1) = \mu_1$. For each $k = 2, ..., |\mathcal{S}|$, define*

$$\mathcal{M}_k = \left\{s \in \left\{\mu_1, \ldots, \mu_{k-1}\right\} : s \succ_{\mu_k} \mu_k \text{ and } \phi(s) = s\right\}. \tag{5}$$

   Then

$$\phi(\mu_k) = \begin{cases} \mu_k & \text{if } \mathcal{M}_k = \varnothing \\ s \in \mathcal{M}_k\text{: } \nexists z \in \mathcal{M}_k \text{ with } z \succ_{\mu_k} s & \text{if } \mathcal{M}_k \neq \varnothing \end{cases}. \tag{6}$$

   (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. *The set of stable states of any two mappings ϕ_1 and ϕ_2 in Φ coincide. That is, $\mathcal{D}_{\phi_1} = \mathcal{D}_{\phi_2} = \mathcal{D}$.*

4. *If, in addition, Assumption 3 holds, then any $\phi \in \Phi$ is "payoff-unique" in the sense that for any two mappings ϕ_1 and ϕ_2 in Φ, $\phi_1(s) \sim \phi_2(s)$ for all $s \in \mathcal{S}$.*

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

$$\text{if } 1 \le j < l \le |\mathcal{S}|\text{, then } \mu_l \not\succ_{\mu_j} \mu_j. \tag{7}$$

The construction is by induction. Assumption 2(a) implies that for any nonempty collection of states $\mathcal{Q} \subset \mathcal{S}$, there exists $z \in \mathcal{Q}$ such that for any $x \in \mathcal{Q}$, $x \not\succ_z z$. Applying this results to $\mathcal{S}$, we obtain μ_1. Now, suppose we have defined μ_j for all $j \leq k-1$, where $k \leq |\mathcal{S}|$. Then, applying the same result to the collection of states $\mathcal{S} \setminus \{\mu_1, \ldots, \mu_{k-1}\}$, we conclude that there exists μ_k satisfying (7) for each k.

The second step is to construct a candidate mapping $\phi : \mathcal{S} \to \mathcal{S}$. 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 $\mathcal{M}_k$ as in (5). This is the subset of states where ϕ has already been defined, which satisfy $\phi(s) = s$ and 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 nonempty, then take $\phi(\mu_k) = z \in \mathcal{M}_k$ such that

$$s \not\succ_{\mu_k} z \text{ for any } s \in \mathcal{M}_k \tag{8}$$

(such state z exists because we can apply Assumption 2(b) to $\mathcal{M}_k$). Proceeding inductively for all $2 \leq k \leq |\mathcal{S}|$, we obtain ϕ as in (6).

To complete the proof, we need to verify that mapping ϕ in (6) satisfies Axioms 1–3. This is straightforward for Axioms 1 and 2. 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 the latter case, $\phi(\mu_k) \succ_{\mu_k} \mu_k$ and $\phi(\phi(\mu_k)) = \phi(\mu_k)$ by (5). To check Axiom 3, 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 (7), implies that $z \in \{\mu_1, \ldots, \mu_{k-1}\}$. But the last condition, $z \succ_{\mu_k} \phi(\mu_k)$, now contradicts (8). This means that such z does not exist, and therefore Axiom 3 is satisfied.

(Part 2) This statement is equivalent to the following: if, given sequence $\{\mu_1, ..., \mu_{|\mathcal{S}|}\}$ with the property (7), $\phi(\mu_k)$ is not given by (6) for some k, then ϕ does not satisfy Axioms 1–3. Suppose, to obtain a contradiction, that $\phi(\mu_k)$ is not given by (6) at $k = 1$. Then by the contradiction hypothesis $\phi(\mu_1) \neq \mu_1$, so $\phi(\mu_1) = \mu_l$ for $l > 1$. In this case, ϕ does not satisfy Axiom 1, because $\mu_l \not\succ_{\mu_1} \mu_1$ by (7), yielding a contradiction. Next, again to obtain a contradiction, suppose that $\phi(\mu_k)$ is not given by (6) for the first time at $k > 1$. Then $\mathcal{M}_k$ in (5) is well-defined, so either $\mathcal{M}_k = \varnothing$ or $\mathcal{M}_k \neq \varnothing$. If $\mathcal{M}_k = \varnothing$, the contradiction hypothesis implies that $\phi(\mu_k) \neq \mu_k$. Then, Axioms 1 and 2 imply $\phi(\mu_k) \succ_{\mu_k} \mu_k$ and $\phi(\phi(\mu_k)) = \phi(\mu_k)$. Since $\mathcal{M}_k = \varnothing$, we must have that $\phi(\mu_k) = \mu_l$ for $l > k$, but in this case $\phi(\mu_k) \succ_{\mu_k} \mu_k$ contradicts (7). This contradiction implies that ϕ violates either Axiom 1 or 2 (or both). If $\mathcal{M}_k \neq \varnothing$ and $\phi(\mu_k) = \mu_l$ for $l > k$, then Axiom 1 is violated. If $\mathcal{M}_k \neq \varnothing$ and $\phi(\mu_k) = \mu_k$, then ϕ violates Axiom 3 (to see this, take any $z \in \mathcal{M}_k \neq \varnothing$ and observe that $z \succ_{\mu_k} \mu_k$ and $\phi(z) = z$). Therefore, when $\mathcal{M}_k \neq \varnothing$, we have $\phi(\mu_k) \in \mathcal{M}_k$. Finally, $\phi(\mu_k)$ will not be given by (6) if there exists some $s \in \mathcal{M}_k$ such that $s \succ_{\mu_k} \phi(\mu_k)$. But in this case ϕ violates Axiom 3 (since

$s \succ_{\mu_k} \phi(\mu_k)$, $s \succ_{\mu_k} \mu_k$, and $\phi(s) = s$). This shows that any mapping ϕ that is not given by (6) violates one of Axioms 1–3. and completes the proof of part 2.

(Part 3) Suppose, to obtain a contradiction, that $\mathcal{D}_{\phi_1} \neq \mathcal{D}_{\phi_2}$. Then $\exists k : 1 \leq k \leq |\mathcal{S}|$ such that $\mu_j \in \mathcal{D}_{\phi_1} \Leftrightarrow \mu_j \in \mathcal{D}_{\phi_2}$ for all $j < k$, but either $\mu_k \in \mathcal{D}_{\phi_1}$ and $\mu_k \notin \mathcal{D}_{\phi_2}$ or $\mu_k \notin \mathcal{D}_{\phi_1}$ and $\mu_k \in \mathcal{D}_{\phi_2}$. Without loss of generality, assume that $\mu_k \in \mathcal{D}_{\phi_1}$ and $\mu_k \notin \mathcal{D}_{\phi_2}$. Then (6) implies that $\phi_2(\mu_k) = \mu_l$ for some $l < k$. Applying Axioms 1 and 2 to mapping ϕ_2, we obtain $\mu_l \succ_{\mu_k} \mu_k$ and $\phi_2(\mu_l) = \mu_l$. The latter implies that $\mu_l \in \mathcal{D}_{\phi_2}$. Since, by hypothesis, $\mu_j \in \mathcal{D}_{\phi_1} \Leftrightarrow \mu_j \in \mathcal{D}_{\phi_2}$ for all $j < k$, we have $\mu_l \in \mathcal{D}_{\phi_1}$. Therefore, $\mu_l \succ_{\mu_k} \mu_k$, $\mu_l \succ_{\mu_k} \phi_1(\mu_k)$ (because $\phi_1(\mu_k) = \mu_k$), and $\phi_1(\mu_l) = \mu_l$, but this violates Axiom 3 for mapping ϕ_1 and establishes the desired result.

(Part 4) Suppose Assumption 3 holds. Suppose, to obtain a contradiction, that ϕ_1 and ϕ_2 are two non-equivalent mappings that satisfy Axioms 1–3, that is, there exists some state s such that $\phi_1(s) \nsim \phi_2(s)$. Part 3 of this Theorem (that $\mathcal{D}_{\phi_1} = \mathcal{D}_{\phi_2}$) implies that $\phi_1(s) = s$ if and only if $\phi_2(s) = s$; since $\phi_1(s) \nsim \phi_2(s)$, we must have that $\phi_1(s) \neq s \neq \phi_2(s)$. Axiom 1 then 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 suppose that the former is the case. 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 equality holds because $\phi_1(s)$ is a ϕ_1-stable state by Axiom 2, and by part 3 of this Theorem, it is also a ϕ_2-stable state). This implies that we can apply Axiom 3 to ϕ_2 and derive the conclusion that $\phi_2(s) \neq y$. 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 set of mappings Φ that satisfy Axioms 1–3. Intuitively, Assumption 2(a) ensures that there exists some state $\mu_1 \in \mathcal{S}$, such that there does not exist another $s \in \mathcal{S}$ with $s \succ_{\mu_1} \mu_1$. Taking μ_1 as base, we order the states as $\left\{\mu_1, ..., \mu_{|\mathcal{S}|}\right\}$ so that (7) is satisfied and then recursively construct the set of states $\mathcal{M}_k \subset \mathcal{S}$, $k = 2, ..., |\mathcal{S}|$, that includes stable states that are preferred to state μ_k (that is, states s such that $\phi(s) = s$ and $s \succ_{\mu_k} \mu_k$). When the set $\mathcal{M}_k$ is empty, then there exists no stable state that is preferred to μ_k (in μ_k) by members of a winning coalition. In this case, we have $\phi(\mu_k) = \mu_k$. When $\mathcal{M}_k$ is nonempty, there exists such a stable state and thus $\phi(\mu_k) = s$ for some such s. 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 of winning coalitions in μ_k favoring a transition to another *stable* state (i.e., by the fact that $\mathcal{M}_k = \varnothing$). We will see that this insight plays an important role in the applications in Section 6.

Part 3 of Theorem 1 shows that the set of stable states $\mathcal{D}$ does not depend on the specific ϕ chosen from Φ. For different ϕ's in Φ (when Φ is not a singleton), the stable state corresponding to the same initial state may differ, but the ranges of these mappings are the same. These ranges and the set of stable states $\mathcal{D}$ are uniquely determined by preferences and the structure of winning coalitions.[7] Finally, part 4 shows that when Assumption 3 holds, any stable states resulting from an initial state must be equivalent. In other words, if $s_1 = \phi_1(s_0)$ and $s_2 = \phi(s_0)$, then s_1 and s_2 might differ in terms of the structure of winning coalitions, but they must give the same payoff to all individuals.

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", i.e., for $t \geq T$ for some finite T). The analysis of the dynamic game in the next section substantiates this interpretation.

4 Noncooperative Foundations of Dynamically Stable States

We now describe the extensive-form game capturing dynamic interactions in the environment of Section 2 and characterize the MPE of this game. The main result is the equivalence between the MPE of this game and the axiomatic characterization in Theorem 1.

This game specifies: (1) a protocol for a sequence of agenda-setters and proposals in each state; and (2) a protocol for voting over proposals. Voting is sequential and is described below (the exact sequence in which votes are cast will not matter). We represent the protocol for agenda-setting using a sequence of mappings, $\{\pi_s\}_{s\in\mathcal{S}}$, and refer to it simply as a *protocol.* Let K_s be a natural number for each $s \in \mathcal{S}$. Then, π_s is defined as a mapping

$$\pi_s : \{1, \ldots, K_s\} \to \mathcal{I} \cup \mathcal{S}$$

for each state $s \in \mathcal{S}$. Thus each π_s specifies a finite sequence of elements from $\mathcal{I} \cup \mathcal{S}$, where K_s is the length of sequence for state s and determines the sequence of agenda-setters and proposals. In particular, if $\pi_s(k) \in \mathcal{I}$, then it denotes an agenda-setter who will make a proposal from the set of states $\mathcal{S}$. Alternatively, if $\pi_s(k) \in \mathcal{S}$, then it directly corresponds to an exogenously-specified proposal over which individuals vote. Therefore, the extensive-form game is general

[7] In Appendix C, we relate the set $\mathcal{D}$ to two concepts from cooperative game theory, von Neumann-Morgenstern's stable set and Chwe's largest consistent set. Even though these concepts generally differ, we show that under Assumptions 1 and 2, both sets coincide with $\mathcal{D}$. This is an intriguing result, though it does not obviate the need for our axiomatic characterization in this section or the noncooperative analysis in the next section, since our main focus is on the mappings ϕ, which determine the axiomatically or dynamically stable states as a function of the initial state. Von Neumann-Morgenstern's stable set and Chwe's largest consistent set are silent on this relationship.

enough to include both proposals for a change to a new state initiated by agenda-setters or exogenous proposals. We make the following assumption on $\{\pi_s\}_{s\in\mathcal{S}}$:

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

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

(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 the "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 (unless a proposal is accepted earlier).

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). Subsequently (for $t \geq 1$), the timing of events is as follows:

1. Period t begins with state s_{t-1} inherited from the previous period.

2. For $k = 1, \ldots, K_{s_{t-1}}$, the kth 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 $P_{k,t}$ 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 $Y_{k,t} \in \mathcal{W}_{s_{t-1}}$, then alternative $P_{k,t}$ is accepted, otherwise (that is, if $Y_{k,t} \notin \mathcal{W}_{s_{t-1}}$), 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, then we transition to state $s_t = P_{k,t}$, and the period ends. If $P_{k,t}$ is rejected or if there is no voting because $P_{k,t} = s_{t-1}$ and $k < K_{s_{t-1}}$, then the game moves to step 2 with k increased by 1; if $k = K_{s_{t-1}}$, the next state is $s_t = s_{t-1}$, and the period ends.

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

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

$$u_i(t) = \begin{cases} w_i(s_t) & \text{if } s_t = s_{t-1} \\ 0 & \text{if } s_t \neq s_{t-1} \end{cases} \tag{9}$$

for each $i \in \mathcal{I}$. In other words, in the period in which a transition occurs, each individual receives zero payoff. In all other periods, each individual i receives the payoff $w_i(s_t)$ as a function of the current state s_t. 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 a pure-strategy MPE. Since the game is infinitely-repeated and we will take β to be large, this one-period "transaction cost" has little effect on discounted payoffs. In particular, once (and if) a dynamically stable state s is reached, individuals will receive $w_i(s)$ at each date thereafter.[8] Examples 3 and 4 in Appendix C demonstrate that if the transaction cost is removed from (9), a (pure-strategy) equilibrium may fail to exist or may induce cycles along the equilibrium path.

A MPE is defined in the standard fashion as a subgame perfect equilibrium (SPE) where strategies are only functions of "payoff-relevant states." Here payoff-relevant states are different from the states $s \in \mathcal{S}$ described above, since the order in which proposals have been made within a given period are also payoff relevant for the continuation game. Since the notion of MPE is familiar, we do not provide a formal definition. For completeness, such a definition is provided in Appendix C. In what follows, we will use the terms MPE and equilibrium interchangeably. We next define dynamically stable states.

Definition 2 ***(Dynamically Stable States)*** *State $s^\infty \in \mathcal{S}$ is a* ***dynamically stable state*** *if there exist an initial state $s \in \mathcal{S}$, a set of protocols $\{\pi_s\}_{s \in \mathcal{S}}$, a MPE strategy profile σ, 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 some finite time T and is repeated thereafter. Our objective is to determine whether dynamically stable states exist in the dynamic game described above and to characterize them as a function of the initial state $s_0 \in \mathcal{S}$. We also establish the equivalence between dynamically and axiomatically stable states characterized in the previous section. We first introduce a slightly stronger version of Assumption 2(b).

Assumption 2(b)* *For any sequence of states $s, s_1, \ldots, s_k$ in $\mathcal{S}$ with $s_j \succ_s s$ for $1 \leq j \leq k$ and $s_j \nsim s_l$ for $1 \leq j < l \leq k$,*

$$s_{j+1} \succeq_s s_j \text{ for all } 1 \leq j \leq k-1 \implies s_1 \not\succeq_s s_k.$$

Moreover, if for $x, y, s \in \mathcal{S}$ we have $x \succ_s s$ and $y \not\succ_s s$, then $y \not\succ_s x$.

[8] Various different alternative game forms also lead to the same results. We chose to present this one because it appears to be the simplest one to describe and encompasses 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 to have a move), which is a desirable feature, since otherwise some transitions would be ruled out by the game form.

In addition to cycles of the form $y \succ_s x$, $z \succ_s y$, $x \succ_s z$ (which are ruled out by Assumption 2(b)), this assumption rules out cycles of the form $y \succeq_s x$, $z \succeq_s y$, $x \succeq_s z$, unless the states x, y, and z are payoff-equivalent. It also imposes the technical requirement that when $x \succ_s s$ and $y \not\succ_s s$, then $y \not\succ_s x$. Both requirements of this assumption are relatively mild.

The main result of the paper is summarized in the following theorem.

Theorem 2 ***(Characterization of Dynamically Stable States)*** *Suppose that Assumptions 1, 2(a,b) and 4 hold. Then there exists $\beta_0 \in (0,1)$ such if for all $\beta > \beta_0$, we have the following results.*

1. *For any $\phi \in \Phi$ there exists a set of protocols $\{\pi_s\}_{s \in \mathcal{S}}$ and a pure-strategy 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 protocols $\{\pi_s\}_{s \in \mathcal{S}}$ there exists a pure-strategy MPE. Any such MPE σ has the property that for any initial state $s_0 \in \mathcal{S}$, $s_t = s^\infty$ for all $t \geq 1$. Moreover, there exists $\phi \in \Phi$ 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 protocols $\{\pi_s\}_{s \in \mathcal{S}}$, any pure-strategy MPE σ induces $s_t \sim \phi(s_0)$ for all $t \geq 1$, where $\phi \in \Phi$.*

Proof. See Appendix A. ■

Parts 1 and 2 of Theorem 2 state that the set of dynamically stable states and the set of stable states $\mathcal{D}$ defined by axiomatic characterization in Theorem 1 coincide; any mapping $\phi \in \Phi$ (satisfying Axioms 1–3) is the outcome of a pure-strategy MPE and any such MPE implements the outcome of some $\phi \in \Phi$. This theorem therefore establishes the equivalence of axiomatic and dynamic characterizations. An important implication is that the recursive characterization of axiomatically stable states in (6) applies exactly to dynamically stable states.

The equivalence of 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 $\beta > \beta_0$, they care more about payoffs in the ultimate state than the payoffs along the transitional states. Consequently, winning coalitions are not willing to move to a state that is not (axiomatically) stable according to Theorem 1; 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 to 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. We start with a simple definition.

Definition 3 ***(Myopic Stability)*** *A state $s^m \in \mathcal{S}$ is* ***myopically stable*** *if there does not exist $s \in \mathcal{S}$ with $s \succ_{s^m} s^m$.*

Myopic stability would apply if individuals made choices only considering their 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 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, in particular, of equation (5). Its proof is omitted.

Corollary 1

1. *State $s^\infty \in S$ 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. *The set of myopically stable states is a subset of $\mathcal{D}$ (the set of axiomatically and dynamically stable states). In particular, a myopically stable state s^m is a stable state, but a stable state s^∞ is not necessarily myopically stable.*

The final part of the corollary implies that s^∞ may be stable even if it is not myopically stable (recall Example 1). In particular, there may exist a state s such that $s \succ_{s^\infty} s^\infty$; but s^∞ may still be stable because $s \neq \phi(s)$ and leads to some other state s', which is not preferred by a winning coalition in s^∞ (if we had $s' = \phi(s) \succ_{s^\infty} s^\infty$, then s^∞ would not be a stable state). Another direct implication of this corollary is that forward-looking behavior enlarges the set of stable states.

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

Definition 4 ***(Inefficiency)*** *State $s \in S$ is* ***(strictly) Pareto inefficient*** *if $\mathcal{W}_s \neq \varnothing$ and there exists a state $s' \in \mathcal{S}$ such that $w_i(s') > w_i(s)$ for all $i \in \mathcal{I}$.*

State $s \in S$ is ***(strictly) winning coalition inefficient*** *if there exists state $s' \in \mathcal{S}$ such that $s' \succ_s s$.*

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

Corollary 2 *1. A stable state $s^\infty \in \mathcal{S}$ can be 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 $s \in \mathcal{S}$ such that $s \succ_{s^\infty} s^\infty$. ■

5 Ordered States

Theorems 1 and 2 provide a complete characterization of axiomatically and dynamically stable states as a function of the initial state $s_0 \in \mathcal{S}$ provided that Assumptions 1 and 2 are satisfied. While the former is a very natural assumption and easy to check, Assumption 2 may be somewhat more difficult to verify. In this section, we show that when the set of states $\mathcal{S}$ admits a (linear) order according to which individual (stage) payoffs satisfy single-crossing or single-peakedness properties (and the set of winning coalitions $\{\mathcal{W}_s\}_{s\in\mathcal{S}}$ satisfies some natural additional conditions), Assumption 2 is satisfied. This result enables more straightforward application of our main theorems in a wide variety of circumstances.

In a number of applications, the set of states $\mathcal{S}$ has a natural order, so that any two states x and y can be ranked (e.g., either x is "greater than" or "less than" y). When such an order exists, we can take, without loss of any generality, $\mathcal{S}$ to be a subset of $\mathbb{R}$. Similarly, let $\mathcal{I} \subset \mathbb{R}$, which is also without loss of any generality. Given these orders on the set of states and the set of individuals, we introduce certain well-known restrictions on preferences.[9] All of the following restrictions and definitions refer to stage payoffs and are thus easy to verify.

Definition 5 ***(Single-Crossing)*** *Given $\mathcal{I} \subset \mathbb{R}$, $\mathcal{S} \subset \mathbb{R}$, and $\{w_i(s)\}_{i\in\mathcal{I},s\in\mathcal{S}}$, the* ***single-crossing condition (SC)*** *holds if, for any $i, j \in \mathcal{I}$ and $x, y \in \mathcal{S}$ such that $i < j$ and $x < y$, $w_i(y) > w_i(x)$ implies $w_j(y) > w_j(x)$ and $w_j(y) < w_j(x)$ implies $w_i(y) < w_i(x)$.*

Definition 6 ***(Single-Peakedness)*** *Given $\mathcal{I} \subset \mathbb{R}$, $\mathcal{S} \subset \mathbb{R}$, and $\{w_i(s)\}_{i\in\mathcal{I},s\in\mathcal{S}}$, preferences are* ***single-peaked (SP)*** *if for any $i \in \mathcal{I}$ there exists state x such that for any $y, z \in \mathcal{S}$, $y < z \le x$ or $x \ge z > y$ implies $w_i(y) \le w_i(z)$.*

We next introduce a generalization of the notion of the "median voter" to more general political institutions (e.g., those involving supermajority rules within the society or a club).

[9] See Barbera and Moreno (2008) for the connection between these notions.

Definition 7 ***(Quasi-Median Voter)*** *Given $\mathcal{I} \subset \mathbb{R}$, $\mathcal{S} \subset \mathbb{R}$, and $\{\mathcal{W}_s\}_{s\in\mathcal{S}}$, player $i \in \mathcal{I}$ is a* ***quasi-median voter*** *(in state s) if for any $X \in \mathcal{W}_s$ such that $X = \{j \in \mathcal{I} : a \leq j \leq b\}$ for some $a, b \in \mathbb{R}$ we have $i \in X$.*

Denote the set of quasi-median voters in state s by M_s. Theorem 3 shows that it is nonempty (provided that Assumption 1 is satisfied).

Definition 8 ***(Monotonic Median Voter Property)*** *Given $\mathcal{I} \subset \mathbb{R}$ and $\mathcal{S} \subset \mathbb{R}$, the sets of winning coalitions $\{\mathcal{W}_s\}_{s\in\mathcal{S}}$ has* ***monotonic median voter property*** *if for each $x, y \in \mathcal{S}$ satisfying $x < y$ there exist $i \in M_x$, $j \in M_y$ such that $i \leq j$.*

The last definition is general enough to encompass majority and supermajority voting as well as these voting rules that apply for a subset of players (such as club members or those that are part of a limited franchise). Finally, we also impose the following weak genericity assumption.

Assumption 5 ***(Weak Genericity)*** *Preferences $\{w_i(s)\}_{i\in\mathcal{I},s\in\mathcal{S}}$ and the set of winning coalitions $\{\mathcal{W}_s\}_{s\in\mathcal{S}}$ are such that for any $x, y, z \in \mathcal{S}$, $x \succeq_z y$ implies $x \succ_z y$ or $x \sim y$.*

Assumption 5 is satisfied if no player is indifferent between any two states (though it does not rule out such indifferences). The main results of this section are presented in the following theorem.

Theorem 3 ***(Characterization with Ordered States)*** *For any $\mathcal{I} \subset \mathbb{R}$, $\mathcal{S} \subset \mathbb{R}$, preferences $\{w_i(s)\}_{i\in\mathcal{I},s\in\mathcal{S}}$, and winning coalition $\{\mathcal{W}_s\}_{s\in\mathcal{S}}$ satisfying Assumption 1, we have that:*

1. *If single-crossing condition and monotonic median voter property hold, then Assumption 2(a,b) is satisfied and thus Theorem 1 applies.*

2. *If preferences are single-peaked and for any $x, y \in \mathcal{S}$ and any $X \in \mathcal{W}_x$, $Y \in \mathcal{W}_y$ we have $X \cap Y \neq \varnothing$, then Assumption 2(a,b) is satisfied and thus Theorem 1 applies.*

3. *If in either part 1 or 2, Assumption 5 also holds, then Assumption 2(b)* is also satisfied and thus Theorem 2 applies.*

Proof. See Appendix A. ■

The conditions of Theorem 3 are relatively straightforward to verify and can be applied in a wide variety of applications and examples. Notice that part 2 of Theorem 3 requires a stronger condition than the monotonic median voter property. It can be verified that this condition implies the monotonic median voter property, so part 1 of the theorem continues to be true under the hypothesis of part 2. However, the converse is not true.[10]

[10] In particular, consider the following environment: there are two states, $x < y$, and two voters, $i < j$.

6 Applications

We now illustrate how the characterization results provided in Theorems 1 and 2 can be applied in a variety of situations, including a number of political economy environments considered in the literature. We show that in many of these environments we can simply appeal to Theorem 3. Nevertheless, we will also see that the conditions in Theorem 3 are more restrictive than those stipulated in Theorems 1 and 2. Thus, when Theorem 3 does not apply, Theorems 1 and 2 may still be applied directly.

6.1 Voting in Clubs

Let us return to Example 2. The society consists of N individuals, so $\mathcal{I} = \{1, \ldots, N\}$. Following Roberts (1999), suppose that there are N states, of the form $s_k = \{1, \ldots, k\}$ for $1 \leq k \leq N$. Roberts (1999) imposes the following strict increasing differences condition:

$$\text{for all } l > k \text{ and } j > i, \;\; w_j(s_l) - w_j(s_k) > w_i(s_l) - w_i(s_k), \tag{10}$$

and considers two voting rules: majority voting within a club (where in club s_k one needs more than $k/2$ votes for a change in club size) or median voter rule (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). These two voting rules lead to corresponding equilibrium notions, which Roberts calls Markov Voting Equilibrium and Median Voter Equilibrium, respectively. He establishes the existence of mixed-strategy equilibria with both notions and shows that they both lead to the same set of stable clubs.

It is straightforward to verify that Roberts's model and his two voting rules are special cases of the general voting rules allowed in our framework. In particular, let us first weaken Roberts's strict increasing differences property to single-crossing, in particular, let us assume that

$$\begin{aligned} \text{for all } l > k \text{ and } j > i, \qquad & w_i(s_l) > w_i(s_k) \implies w_j(s_l) > w_j(s_k), \text{ and} \\ & w_j(s_k) > w_j(s_l) \implies w_i(s_k) > w_i(s_l). \end{aligned} \tag{11}$$

Clearly, (10) implies (11) (but not vice versa). In addition, Roberts's two voting rules can be represented by the following sets of winning coalitions:

$$\mathcal{W}_{s_k}^{maj} = \{X \in \mathcal{C} : |X \cap s_k| > k/2\}, \text{ and}$$

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

Preferences are such that $w_i(x) < w_i(y)$, but $w_j(x) > w_j(y)$. These preferences are single-peaked (so are any preferences with two states). Suppose, in addition, that $\mathcal{W}_x = \{\{i\}, \{i,j\}\}$, and $\mathcal{W}_y = \{\{j\}, \{i,j\}\}$. In that case, i and j are quasi-median voters in states x and y, respectively, and thus monotonic median voter property holds. However, Assumption 2(a) is violated for $\{x, y\}$ (we have $y \succ_x x$ and $x \succ_y y$).

Clearly, both $\left\{\mathcal{W}_{s_k}^{maj}\right\}_{k=1}^{N}$ and $\left\{\mathcal{W}_{s_k}^{med}\right\}_{k=1}^{N}$ satisfy Assumption 1 as well as the monotonic median voter property in Definition 8. Let us also assume that Assumption 5 holds. In this case, this can be guaranteed by assuming that $w_i(s) \neq w_i(s')$ for any $i \in \mathcal{I}$ and any $s, s' \in \mathcal{S}$ (though a weaker condition would also be sufficient). Then, it is clear that Theorem 3 from the previous section applies to Roberts's model and establishes the existence of a pure-strategy MPE and characterizes the structure of stable clubs.

It can also be verified that Theorem 3 applies with considerably more general voting rules. For example, we could allow a different degree of supermajority rule in each club. The following set of winning coalitions nests various majority and supermajority rules: for each k, let the degree of supermajority in club s_k be l_k where $k/2 < l_k \leq k$ and define the set a winning coalitions as:

$$\mathcal{W}_{s_k}^{l_k} = \{X \in \mathcal{C} : |X \cap s_k| \geq l\}$$

Then, a relatively straightforward application of Theorem 3 establishes the following proposition (for completeness, the proof is provided in Appendix C).

Proposition 1 *In the voting in clubs model, with winning coalitions given by either $\mathcal{W}_{s_k}^{maj}$, $\mathcal{W}_{s_k}^{med}$, or $\mathcal{W}_{s_k}^{l_k}$, where $k/2 < l_k \leq k$ for all k, the following results hold.*

(i) The monotonic median voters property in Definition 8 is satisfied.

(ii) Suppose that preferences satisfy (11) *and Assumption 5. Then Assumptions 2(a,b) and 2b* hold and thus the characterization of MPE and stable states in Theorems 1 and 2 applies.*

(iii) Moreover, if only odd-sized clubs are allowed, then in the case of majority or median voter rules Assumption 3 also holds and thus the dynamically stable state (club) is uniquely determined as a function of the initial state (club).

This proposition shows that a sharp characterization of dynamics of clubs and the set of stable clubs can be obtained easily by applying Theorem 3 to Roberts's original model or to various generalizations. Another generalization, not stated in Proposition 1, is to allow for a richer set of clubs. For example, the feasible set of clubs can also be taken to be of the form of $\{k-n, \ldots, k, \ldots, k+n\} \cap \mathcal{I}$ for a fixed n (and different values of k). It is also noteworthy that the approach in Roberts's paper is considerably more difficult and restrictive (though Roberts also establishes the existence of mixed-strategy MPE for any β). Therefore, this application illustrates the usefulness of the general characterization results presented in this paper.

6.2 The Structure of Elite Clubs

In this subsection, we briefly discuss another example of dynamic club formation, which allows a simple explicit characterization. Suppose there are N individuals $1, 2, \ldots, N$ and N states

$s_1, s_2, \ldots, s_N$, where $s_k = \{1, 2, \ldots, k\}$. Preferences are such that for any $n_0 = n_1 < j \leq n_2 < n_3$,

$$w_k(s_{n_0}) = w_k(s_{n_1}) < w_k(s_{n_3}) < w_k(s_{n_2}). \tag{12}$$

These preferences imply that each player k wants to be part of the club, but conditional on being in the club, he prefers to be in a smaller (more "elite") one. In addition, a player is indifferent between two clubs he is not part of. Suppose that decisions are made by a simple majority rule of the club members, so that winning coalitions are given by

$$\mathcal{W}_{s_k} = \{X \in \mathcal{C} : |X \cap s_k| > k/2\}. \tag{13}$$

It is straightforward to verify that this environment satisfies Assumptions 1, 2(a,b), 2b*, and 3.[11] Hence, we can use Theorems 1 and 2 to characterize the set of stable states and the unique outcome mapping. First, notice that state s_1 is stable. This club only includes player 1, who is thus the dictator, and who likes this state best, and thus by Axiom 1 we must have $\phi(s_1) = s_1$. In state s_2, a consensus of players 1 and 2 is needed for a change. But s_2 is the best state for player 2, so $\phi(s_2) = s_2$. In state s_3, the situation is different: state s_2 is stable and is preferred to s_3 by both 1 and 2 (and is the only such state), so $\phi(s_3) = s_2$. Proceeding inductively, we can show that club s_j is stable if and only if $j = 2^n$ for $n \in \mathbb{Z}_+$, and the unique mapping ϕ that satisfies Axioms 1–3 is

$$\phi(s_k) = s_{2^{\lfloor \log_2 k \rfloor}}, \tag{14}$$

where $\lfloor x \rfloor$ denotes the greatest integer less than or equal to $x \in \mathbb{R}$.

The following proposition summarizes the above discussion.

Proposition 2 *In the elite club example considered above with preferences given by* (12) *and set of winning coalitions given by* (13)*, the following results hold.*

1. *Assumptions 1, 2(a,b), 2b*, and 3 hold.*

2. *If, instead of* (12)*, for $n_0 < n_1 < k \leq n_2 < n_3$ we have $w_k(s_{n_0}) < w_k(s_{n_1}) < w_k(s_{n_3}) < w_k(s_{n_2})$, then single-crossing condition is satisfied (and monotonic median voter property is always satisfied in this example).*

3. *Club s_k is stable if and only if $k = 2^n$ for $n \in \mathbb{Z}^+$.*

[11] This is formally shown in Appendix C. Alternatively, one could consider a slight variation where a player who does not belong to either of any two clubs prefers the larger of the two. In this case, Theorem 3 can also be applied. In particular, with this variation, the single-crossing condition is satisfied (if $w_i(s_y) > w_i(s_x)$ for $y > x$ and $j > i$, then $i \notin x$ and thus, $j \notin x$, and $w_j(s_y) > w_j(s_x)$; conversely, $w_j(s_y) < w_j(s_x)$ means $j \in s_y$, thus $i \in s_y$, and therefore $w_i(s_y) < w_i(s_x)$). The monotonic median voter condition holds as well (one can choose quasi-median voter in state s_j to be $\lfloor (j+1)/2 \rfloor \in M_{s_j}$; this sequence is weakly increasing in j).

4. The unique mapping ϕ that satisfies Axioms 1–3 *is given by* (14).

Proof. See Appendix C. ■

6.3 Stable Voting Rules and Constitutions

Another interesting model that can be analyzed using Theorem 3 is Barbera and Jackson's (2004) model of self-stable constitutions. In addition, our analysis shows how more farsighted decision-makers can be easily incorporated into Barbera and Jackson's model.

Motivated by Barbera and Jackson's model, let us introduce a somewhat more general framework. The society takes the form of $\mathcal{I} = \{1, \ldots, N\}$ and each state now directly corresponds to a "constitution" represented by a pair (a, b), where a and b are integers between 1 and N. The utility from being in state (a, b) is fully determined by a, so that each player i receives utility

$$w_i \left[(a, b)\right] = w_i \left(a\right). \tag{15}$$

In contrast, the set of winning coalitions needed to change the state is determined by $b \in \mathbb{Z}_+$:

$$\mathcal{W}_{(a.b)} = \{X \in \mathcal{C} : |X| \geq b\} \tag{16}$$

(so b may be interpreted as the degree of supermajority).

In Barbera and Jackson's model, individuals differ according to the probability with which they will support a proposal for a specific reform away from the status quo. The parameter a determines the (super)majority necessary for implementing the reform. The parameter b, on the other hand, is the (super)majority necessary (before individual preferences are realized) for changing the voting rule a. Expected utility is calculated before these preferences are realized and defines $w_i \left[(a, b)\right]$. Ranking individuals according to the probability with which they will support the reform, Barbera and Jackson show that individual preferences satisfy (strict) single-crossing and are (weakly) single-peaked.

For our analysis here, let us consider any situation in which preferences and winning coalitions satisfy (15) and (16). It turns out to be convenient to reorder all pairs (a, b) on the real line as follows: if (a, b) and (a', b') satisfy $a < a'$, then (a, b) is located on the left of (a', b'), and we write $(a, b) < (a', b')$; the ordering of states with the same a is unimportant. Suppose that $w_i (a)$, and thus $w_i \left[(a, b)\right]$, satisfies the single-crossing condition in Definition 5. This enables us to apply Theorem 3 to any problem that can be cast in these terms, including the original Barbera and Jackson model.

Let us next follow Barbera and Jackson in distinguishing between two cases. In the case of *constitutions*, any combination (a, b) is allowed, while in the case of *voting rules*, only the subset of states where $a = b$ is considered (then $a = b$ is the voting rule); in both cases it is natural to assume $b > N/2$. Barbera and Jackson call a voting rule or a constitution (a, b) self-stable if there is no alternative voting rule (a', b') with $a' = b'$ (or, respectively, constitution (a', b')) such that (a', b') is preferred to (a, b) by at least b players. Clearly, this stability notion is equivalent to our notion of myopic stability. Given Corollary 1, it is not surprising that when we allow players to be farsighted, the set of stable states may be enlarged.[12]

The following proposition shows summarizes this discussion.

Proposition 3 *Consider the above-described environment and assume that preferences satisfy single-crossing condition and Assumption 5 holds. Then:*

1. *Assumptions 1, 2(a,b) and 2b* are satisfied.*
2. *There exist mappings ϕ_v for the case of voting rules ($a = b$) and ϕ_c for the case of constitutions that satisfy Axioms 1–3.*
3. *In the case of voting rules, the set of self-stable voting rules (in the sense of Barbera and Jackson) coincides with the set of myopically stable states. In particular, any such state (a, b), where $a = b$, satisfies $\phi_v [(a, b)] = (a, b)$. The set of self-stable voting rules is a subset of set of dynamically stable states.*
4. *In the case of constitutions, the set of self-stable constitutions (in the sense of Barbera and Jackson), the set of myopically stable states and the set of dynamically stable states coincide.*

Proof. See Appendix C. ■

6.4 Coalition Formation in Nondemocracies

As mentioned above, Theorems 1 and 2 can be directly applied in situations where the set of states does not admit a (linear) order. We now illustrate one such example using a modification of the game of dynamic coalition formation in Acemoglu, Egorov, and Sonin (2008).

[12] In particular, the set of stable states is enlarged in the case of voting rules, but remains the same in the case of constitutions. This can be seen as follows: suppose constitution (a', b') is preferred to (a, b) by at least b players; without loss of generality, we may assume that (a, b) is Pareto-efficient (otherwise we could pick (a'', b'') which Pareto dominates (a', b') and thus is preferred to (a, b) by these b players). Then these b players also prefer constitution (a', N) to (a, b), since the payoffs are the same. But constitution (a', N) is stable by Axiom 1. Moreover, it is impossible that all N players prefer some other constitution (a'', b'') because (a', b'), and thus (a', N), are Pareto efficient. Hence, if state (a, b) is not myopically stable, it is also not dynamically stable, for the players may move to constitution (a', N), which is dynamically stable and preferred to (a, b) by at least N players.

Suppose that each state determines the ruling coalition in a society and thus the set of states $\mathcal{S}$ coincides with the set of coalitions $\mathcal{C}$. Members of the ruling coalition determine the composition of the ruling coalition in the next period. A transition to any coalition in $\mathcal{C}$ is allowed, which highlights that the set of states does not admit a complete order (one could define a partial order over states, though this is not particular useful for the analysis here).[13]

Each agent $i \in \mathcal{I}$ is assigned a positive number γ_i, which we interpret as "political influence" or "political power." For any coalition $X \in \mathcal{C}$, let

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

Suppose that payoffs are given by

$$w_i(X) = \begin{cases} \gamma_i / \gamma_X & \text{if } i \in X \\ 0 & \text{if } i \notin X \end{cases} \tag{17}$$

for any $i \in \mathcal{I}$ and any $X \in \mathcal{C} \equiv \mathcal{S}$.[14] The restriction to (17) here is just for simplicity. Also, take any $\alpha \in [1/2, 1)$ as a measure of the extent of supermajority requirement. Define the set of winning coalitions as

$$\mathcal{W}_X = \left\{ Y \in \mathcal{C} : \sum\nolimits_{j \in Y \cap X} \gamma_j > \alpha \sum\nolimits_{j \in X} \gamma_j \right\}. \tag{18}$$

Clearly, this corresponds to weighted α-majority voting among members of the incumbent coalition X (with $\alpha = 1/2$ corresponding to simple majority). In addition, suppose that the following simple *genericity* assumption holds:

$$\gamma_X = \gamma_Y \text{ only if } X = Y. \tag{19}$$

The following proposition can now be established.

Proposition 4 *Consider the environment in Acemoglu, Egorov, and Sonin (2008). Then:*

1. *Assumptions 1, 2(a,b), 3 are satisfied, so that Theorem 1 applies and characterizes the axiomatically stable states.*

[13] In Acemoglu, Egorov and Sonin (2008), not all transitions were allowed. In particular, there we focused on a game 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 by other means). Moreover, we assumed that payoffs were realized at the end of the game. Appendix B discusses how the current framework can be generalized so that there are limits on the feasible set of transitions.

[14] This is a special case of the payoff structure in Acemoglu, Egorov and Sonin (2008), where we allowed for any payoff function satisfying the following three properties: (1) if $i \in X$ and $i \notin Y$, then $w_i(X) > w_i(Y)$; (2) 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 (3) $i \notin X$ and $i \notin Y$, then $w_i(X) = w_i(Y)$. The form in (17) is adopted to simplify the discussion here.

2. *Moreover, there exists an arbitrarily small perturbation of payoffs such that Assumption 2(b)* also holds. In this case, Theorem 2 also applies and characterizes the dynamically stable states.*

Proof. See Appendix C. ∎

In Appendix B we introduce the possibility of restrictions on feasible transitions and show how Proposition 4 can be generalized to cover the case of political eliminations considered in Acemoglu, Egorov, and Sonin (2008). We also illustrate how not allowing previously-eliminated players to be part of the ruling coalition affects the results and the structure of stable coalitions.

6.5 Coalition Formation in Democracy

We next briefly discuss how similar issues arise in the context of coalition formation in democracies, for example, in coalition formation in in legislative bargaining.[15]

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 $\varnothing$, and all coalitions are possible states. Since any two parties are sufficient to form a government, we have $\mathcal{W}_{\varnothing} = \mathcal{W}_s = \{\{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$ for all s. 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, $w_3(\varnothing) \leq w_3(\{1,2\}) < w_3(\{1,2,3\}) < w_3(\{1,3\}) < w_3(\{2,3\})$; other payoffs are defined similarly. In this case, it can be verified that $\phi(\varnothing) = \{2,3\}$: 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 (recall footnote 2), the dynamics of coalition formation does 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 $w_2(\varnothing) \leq w_2(\{1,3\}) < w_2(\{1,2,3\}) < w_2(\{2,3\}) < w_2(\{1,2\})$. 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, $\phi(\varnothing) = \{1,2\}$. Nevertheless, whether this will be the case depends on the continuation game after coalition $\{1,2\}$ is formed. Suppose, for example, that after the coalition

[15] See, for example, Baron and Ferejohn (1986), Austen-Smith and Banks (1988), Baron (1991), Jackson and Moselle (2002), and Norman (2002) for models of legislative bargaining. The recent paper by Diermeier and Fong (2008) that studies legislative bargaining as a dynamic game without commitment also raises a range of issues related to our general framework here.

$\{1,2\}$ 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 such a feasible transition, so that we have $\{1,2\} \mapsto \{1\}$ (which naturally presumes that $\mathcal{W}_{\{1,2\}} = \{X \in \mathcal{C} : 1 \in X\}$). Similarly, starting from the coalition $\{2,3\}$, party 2 can also do the same, so that $\mathcal{W}_{\{2,3\}} = \{X \in \mathcal{C} : 2 \in X\}$ and $\{2,3\} \mapsto \{2\}$. However, it 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 thus $\{2\} \mapsto \{1\}$. In this case, the analysis in this paper immediately shows that $\phi(\varnothing) = \{2,3\}$, 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, in the context of coalition formation in democracies, the insight (discussed after Theorem 1 and in Corollary 1) that the instability of states that can be reached from a state s contributes to the stability of state s.

6.6 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, $\mathcal{W}_a = \{X \in \mathcal{C} : E \in X\}$. Suppose that for all $x \in \mathcal{S} \setminus \{a\}$, we have that $\mathcal{I} \setminus \{E\} \in \mathcal{W}_x$, that is, all players except E together 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 that absolutist monarchy, we will eventually end up with democracy. Suppose also that there exists $y \in \mathcal{S}$ such that $w_i(y) > w_i(a)$, 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 then 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, since these will lead to state d and thus $\phi(a) = a$.

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 can no longer block the transition to d.

6.7 Middle Class and Democratization

Let us next consider a variation of the environment discussed in the previous subsection. Suppose again that the initial state is $s_0 = a$, where $\mathcal{W}_a = \{X \in \mathcal{C} : E \in X\}$. 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 $\mathcal{W}_{d1} = \mathcal{W}_{d2} = \{X \in \mathcal{C} : P \in X\}$. Suppose

$$w_E(d2) < w_E(a) < w_E(d1) \text{ and } w_P(a) < w_P(d1) < w_P(d2),$$

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 redistribution. Anticipating this, E will resist democratization.

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

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

The middle class is also opposed to extensive redistribution, so

$$w_M(a) < w_M(d2) < w_M(d1).$$

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). Thus, this example illustrates how the presence of an additional powerful player, such as the middle class, can have a moderating effect on political conflict and enable institutional reform that might otherwise be impossible (see Acemoglu and Robinson, 2006a, for examples in which the middle class may have played such a role in the process of democratization).

6.8 Concessions in Civil War

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, e.g., Fearon, 1996, 2004, Powell, 1998). 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 subsection 6.5, we now have $p \mapsto r$, where r denotes a state in which the rebel group becomes strong and sufficiently influential in domestic politics. Moreover, $\mathcal{W}_p = \{X \in \mathcal{C} : R \in X\}$, and naturally, $w_R(r) > w_R(p)$. If $w_G(r) < w_G(c)$, there will be no peace and $\phi(c) = c$ despite the fact that we may also have $w_G(p) > w_G(c)$. The reasoning for why civil war may continue in this case is similar to that for inefficient inertia 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 $w_G(dp) > w_G(d)$, 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.

6.9 Taxation and Public Good Provision

In many applications preferences are defined over economic allocations, which are themselves determined endogenously as a function of political rules. Our main results can also be applied in such environments. Here we illustrate this by providing an example of taxation and public good provision.

Suppose there are N individuals $1, 2, \ldots, N$ and N states $s_1, s_2, \ldots, s_N$, where $s_k = \{1, 2, \ldots, k\}$. We assume that decisions on transitions are made by an absolute majority rule of individuals who are enfranchised, so that winning coalitions take the form

$$\mathcal{W}_{s_k} = \{X \in \mathcal{C} : |X \cap s_k| > k/2\}.$$

We also assume that the payoff of individual i is given by

$$w_i(s_j) = \mathbb{E}\left[\left(1 - \tau_{s_j}\right) A_i + G_{s_j}\right], \tag{20}$$

where A_i is individual i's productivity (we assume $A_i > A_j$ for $i < j$, so that lower-ranked individuals are more productive), $\mathbb{E}$ denotes the expectations operator, and τ_{s_j} is the tax rate determined when the voting franchises s_j. When an odd number of individuals are allowed to vote, the tax rate is determined by the median. When there is an even number of voters, each of two median voters gets to set the tax rate with equal probability. The expectations in (20) is included because of the uncertainty of the identity of the median voter in this case. Finally, $G_{s_j} = h\left(\sum_{l=1}^{k} \tau_{s_j} A_l\right)$ is the public good provided through taxation, where h is an increasing concave function.

For the single-crossing property, we require that for any $i < j \in \mathcal{I}$ and for any $s_l, s_{l+1} \in \mathcal{S}$,

$$w_j(s_{l+1}) > w_j(s_l) \Rightarrow w_i(s_{l+1}) > w_i(s_l) \text{ and } w_i(s_{l+1}) < w_i(s_l) \Rightarrow w_j(s_{l+1}) < w_j(s_l).$$

Denoting the equilibrium taxes in states s_l and s_{l+1} by $\tau_{s_{l+1}}$ and τ_{s_l}, the following condition is sufficient (but not necessary) to ensure this:

$$\mathbb{E}\left(1-\tau_{s_{l+1}}\right)A_j - \mathbb{E}\left(1-\tau_{s_l}\right)A_j > \mathbb{E}\left(1-\tau_{s_{l+1}}\right)A_i - \mathbb{E}\left(1-\tau_{s_l}\right)A_i,$$

since the equilibrium levels of public goods, G_{s_l} and $G_{s_{l+1}}$, cancel out from both sides. Therefore,

$$\mathbb{E}\tau_{s_{l+1}} > \mathbb{E}\tau_{s_l} \tag{21}$$

is sufficient for single-crossing. Notes that individual i, when determining the tax rate in s_l, would maximize

$$(1-\tau)A_i + h\left(\tau \sum\nolimits_{m=1}^{l} A_m\right).$$

This implies that individual i would choose τ_i such that

$$A_i = h'\left(\tau_i \sum\nolimits_{m=1}^{l} A_m\right)\sum\nolimits_{m=1}^{l} A_m.$$

From the concavity of h it follows that for $i < j$, $\tau_i > \tau_j$. Now consider a switch from s_l to s_{l+1}. Then, with probability 1/2, the tax is set by the same individual (then the tax rate is the same in s_{l+1} as in s_l), and with probability 1/2, by a less productive individual (then the tax rate is greater in s_{l+1} than in s_l). Therefore, (21) holds and we can apply Theorem 3 to characterize the dynamically stable states in this society. More interestingly, these results can also be extended to situations where public goods [taxes] are made available differentially to [imposed on] those who have voting rights (club members).

7 Conclusion

A central feature of collective decision-making in many social situations, such as societies choosing their constitutions or political institutions, or political coalitions, international unions, or private clubs choosing their membership, is that the rules that govern the regulations and procedures 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.

We developed a general framework for a systematic study of this class of problems. We provided both an axiomatic and a noncooperative characterization of stable states and showed that the set of (dynamically) stable states can be computed recursively. This recursive characterization highlights that a particular state s is stable if there does not exist another stable state that makes a winning coalition (in s) better off. An implication of this reasoning is that stable states need not be Pareto efficient; there may exist a state that provides higher payoffs to all individuals, but is itself not stable.

We also showed 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, and Jack and Lagunoff, 2006), models of voting in clubs (such as Roberts, 1999, and Barbera, Maschler, and Shalev, 2001), models of the stability of constitutions (such as Barbera in Jackson, 2004), and models of coalition formation in democracies and nondemocracies. In these cases and in a number of others, we illustrated how models previously studied in the literature are special cases of our framework and how our approach highlights the main economic insights in these diverse environments

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 stochastically vary over time, and to include capital-like state variables so that some subcomponents of the state 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

7.1 Proof of Theorem 2

We start with a lemma about the structure of MPE and SPE in voting games. This lemma plays an important role in the proof of Theorem 2 and is of independent interest in the context of dynamic political economy models. It establishes that there always exists a pure-strategy MPE (and SPE) 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 sufficiently many players (a "winning coalition") will be implemented.

Lemma 1 *Consider the following I-player extensive-form game G_N with perfect information and N stages. In each stage k, one player i_k (this player may be the same for different k's) takes action $a_k \in \{y,n\}$. The payoff vector is given by the mapping $v : \{y,n\}^N \to \{\bar{y},\bar{n}\}$, where $\bar{y}$, $\bar{n} \in \mathbb{R}^I$ are the two possible vectors of payoffs. Suppose that if for some k, $v(a_k = n, a_{-k}) = \bar{y}$, then we also have $v(a_k = y, a_{-k}) = \bar{y}$ (i.e., if for any profile of actions other than that at stage k, a_{-k}, the vote $a_k = n$ leads to $\bar{y}$, then so does $a_k = y$; this ensures that action y does not make outcome $\bar{y}$ less likely). Then the following results hold.*

(i) There exists a pure-strategy MPE (and SPE) in G_N where $a_k = y$ if $v_{i_k}(\bar{y}) > v_{i_k}(\bar{n})$ and $a_k = n$ if $v_{i_k}(\bar{y}) < v_{i_k}(\bar{n})$ (where $v_i(\bar{y})$ and $v_i(\bar{n})$ denote the payoffs of player i under the payoff vectors $\bar{y}$ and $\bar{n}$, respectively).

(ii) Suppose that the set of players $\mathcal{Y} = \{i : v_i(\bar{y}) > v_i(\bar{n})\}$ is large enough, in the sense that $v(a_1,\ldots,a_N) = \bar{y}$ whenever $a_k = y$ for all $i_k \in \mathcal{Y}$. Then in any SPE of G_N, the equilibrium payoff vector will be $\bar{y}$ with probability 1. Similarly, if the set of players $\mathcal{N} = \{i : v_i(\bar{y}) < v_i(\bar{n})\}$ is large enough, so that $v(a_1,\ldots,a_N) = \bar{n}$ whenever $a_k = n$ for all $i_k \in \mathcal{N}$, then in any SPE of G_N, the 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 $G_{N'}$ with perfect information satisfy the requirements above, except that instead of payments at terminal nodes taking two values only, we have that there are two classes of isomorphic subgames, $S_{\bar{y}}$ and $S_{\bar{n}}$, with payoff vectors $\bar{y}$ and $\bar{n}$ respectively. Take any MPE σ and let $\mathcal{Y} = \{i : v_i(S_{\bar{y}}) > v_i(S_{\bar{n}})\}$ and $\mathcal{N} = \{i : v_i(S_{\bar{y}}) < v_i(S_{\bar{n}})\}$, where $v_i(S_{\bar{y}})$ and $v_i(S_{\bar{n}})$ are continuation payoffs of player i. If $v(a_1,\ldots,a_N) = \bar{y}$ whenever $a_k = y$ for all $i_k \in \mathcal{Y}$, then the equilibrium continuation game reached after N stages is $S_{\bar{y}}$ and the expected utility players receive in this MPE is $v(S_{\bar{y}})$. Conversely, if $v(a_1,\ldots,a_N) = \bar{n}$ whenever $a_k = n$ for all $i_k \in \mathcal{N}$, then the equilibrium continuation game reached after N stages is $S_{\bar{n}}$ and the expected utility players receive in this MPE is $v(S_{\bar{n}})$.

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 (this will establish the existence of a pure-strategy MPE and SPE). Consider player i_k such that $v_{i_k}(\bar{y}) > v_{i_k}(\bar{n})$ and suppose that 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 $a'_k = y$ can 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 payoff and thus this player does not have a profitable deviation, which completes the proof of part 1.

(Part 2) We prove this by induction on the number of stages k.

Base: take $k = 1$. Suppose that set $\mathcal{Y}$ is large enough, so that player i_1 choosing action $a_1 = y$ is sufficient for the payoff vector to be $\bar{y}$. To obtain a contradiction, suppose that in a SPE 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 chooses 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 $\mathcal{N}$ is large enough. We have thus proved the base.

Step from $k-1$ to k: suppose that we have proved the result for all $l \leq k-1$; consider the game with k stages. Suppose that set $\mathcal{Y}$ is large enough. Consider two cases.

Case 1: suppose that $v_{i_1}(\bar{y}) > v_{i_1}(\bar{n})$. In this case, if player i_1 in stage 1 takes action $a_1 = y$, then in the subgame starting at stage 2 the following is true: if all players for whom $v_{i_j}(\bar{y}) > v_{i_j}(\bar{n})$ for all $2 \leq j \leq k$ choose $a_j = y$, then the payoff vector will be $\bar{y}$. By induction, any SPE in this subgame will lead to $\bar{y}$ with probability 1. Now if for some SPE of the entire game the payoff vector were $\bar{n}$ with positive probability, then player i_1 could ensure that the payoff vector is $\bar{y}$ with probability 1 by choosing $a_1 = y$ and would thus have a profitable deviation. Therefore, in this case, the payoff vector must be $\bar{y}$ with probability 1.

Case 2: suppose that $v_{i_1}(\bar{y}) \leq v_{i_1}(\bar{n})$. Then, by assumption, if in the subgame starting at stage 2 all players for whom $v_{i_j}(\bar{y}) > v_{i_j}(\bar{n})$ for all $2 \leq j \leq k$ choose $a_j = y$, then the payoff vector will be $\bar{y}$. By induction, for any SPE in any subgame starting at stage 1, the payoff vector is $\bar{y}$ with probability 1. But this implies that the same holds for the entire k-stage game.

The two cases together complete the induction step for the case where $\mathcal{Y}$ is large enough. The case where $\mathcal{N}$ is large enough is analogous. This argument completes the proof of part 2.

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

7.2 Proof of Theorem 2

Proof of Theorem 2 (Part 1) First, suppose that β satisfies the following conditions:

$$\begin{aligned} &\text{for any } i \in \mathcal{I} \text{ and } s, s' \in \mathcal{S}, \\ w_i\left(s'\right) &< w_i\left(s\right) \text{ implies } \beta^{|\mathcal{S}|} > \frac{w_i\left(s'\right)}{w_i\left(s\right)} \text{ and } \frac{1-\beta}{\beta} < \frac{w_i\left(s\right) - w_i\left(s'\right)}{\max_{z\in\mathcal{S}} w_i\left(z\right)}. \end{aligned} \tag{A1}$$

There is a finite number (not more than $|\mathcal{I}| \times |\mathcal{S}| \times (|\mathcal{S}| - 1)$) of conditions in (A1). It is straightforward to verify that there exists $\beta_0 \in (0,1)$ such that for all $\beta > \beta_0$, (A1) holds.

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

$$V_i\left(s,q\right) = \left\{ \begin{array}{cl} \left(1-\beta\right) w_i\left(s\right) & \text{if } s = q \\ 0 & \text{if } s \neq q \end{array} \right\} + \left\{ \begin{array}{cl} \beta w_i\left(\phi\left(q\right)\right) & \text{if } \phi\left(q\right) = q \\ \beta^2 w_i\left(\phi\left(q\right)\right) & \text{if } \phi\left(q\right) \neq q \end{array} \right\}. \tag{A2}$$

(This means that $V_i\left(s,q\right)$ is given by one of the four expressions in (A2) depending on whether $s = q$ and $\phi\left(q\right) = q$). In the equilibrium we construct below, $V_i\left(s,q\right)$ will be the continuation payoff of player i as a function of the current state s and accepted proposal is q. Given the focus on MPE, we drop the time indices.

For each $s \in \mathcal{S}$, take $K_s \geq |\mathcal{S}| - 1$. Take $\pi_s\left(\cdot\right)$ such that $\pi_s\left(K_s\right) = \phi\left(s\right)$ if $\phi\left(s\right) \neq s$; otherwise, take $\pi_s\left(k\right)$ arbitrarily (making sure that Assumption 4 is satisfied). Consider the strategy profile σ^* constructed as follows:

Each player player $i \in \mathcal{I}$ votes for proposal P_k (says *yes*) if and only if:

(i) either $k = K_s$ (we are at the last stage of voting), $P_{K_s} = \phi\left(s\right)$ and $V_i\left(s,\phi\left(s\right)\right) > V_i\left(s,s\right)$;

(ii) or $V_i\left(s,P_k\right) > V_i\left(s,\phi\left(s\right)\right)$.

In addition, if $\pi_s\left(k\right) \in \mathcal{I}$ for some k, this player chooses proposal P_k arbitrarily.

The strategy profile σ^* is Markovian. We will show that it is an MPE in three steps.

First, we will show that under the strategy profile σ^*, there is a transition to $\phi\left(s\right)$ if $\phi\left(s\right) \neq s$ and no transition if $\phi\left(s\right) = s$. If $\phi\left(s\right) \neq s$, then Axiom 1 implies that the set of players for whom $V_i\left(s,\phi\left(s\right)\right) > V_i\left(s,s\right)$ is a winning coalition in s, that is,

$$X_s \equiv \left\{i : w_i\left(\phi\left(s\right)\right) > w_i\left(s\right)\right\} \in \mathcal{W}_s.$$

To see this, observe that (A1) and the fact that $\beta > \beta_0$ imply that $\beta w_i\left(\phi\left(s\right)\right) > w_i\left(s\right)$ for all $i \in X_s$. Therefore, for all $i \in X_s$, we have

$$V_i\left(s,\phi\left(s\right)\right) = \beta w_i\left(\phi\left(s\right)\right) > \left(1-\beta\right) w_i\left(s\right) + \beta^2 w_i\left(\phi\left(s\right)\right) = V_i\left(s,s\right).$$

Next, we can similarly show that there exists no $X'_s \in \mathcal{W}_s$ such that $V_i\left(s,P_k\right) > V_i\left(s,\phi\left(s\right)\right)$ for all $i \in X'_s$, that is, the set of players for whom $V_i\left(s,P_k\right) > V_i\left(s,\phi\left(s\right)\right)$ does not form a winning coalition in s. To obtain a contradiction, suppose there exists such a X'_s. Then since $P_k \neq s$ and $\phi(\phi(s)) = \phi(s)$, we would have that

$$\beta w_i(\phi(P_k)) \geq V_i(s, P_k) > V_i(s, \phi(s)) \geq \beta w_i(\phi(s)) \text{ for all } i \in X'_s,$$

and thus

$$w_i(\phi(P_k)) > w_i(\phi(s)) \text{ for all } i \in X'_s.$$

Then the fact that $X'_s \in \mathcal{W}_s$ implies $\phi(P_k) \succ_s \phi(s)$, which, given that $\phi(s) \succ_s s$, yields $\phi(P_k) \succ_s s$ by Assumption 2(b)*. But $\phi(P_k) \succ_s \phi(s)$, $\phi(P_k) \succ_s s$, and $\phi(P_k) = P_k$ contradicts Axiom 3 and yields a contradiction to our hypothesis that $X'_s \in \mathcal{W}_s$. Therefore, the set of players with $V_i(s, P_k) > V_i(s, \phi(s))$ does not form a winning coalition in s. We have therefore established that under σ^*, $P_{K_s} = \phi(s)$ if $\phi(s) \neq s$ is accepted and all other proposals are rejected.

Second, we verify that given σ^*, continuation payoffs after acceptance of proposal q are given by (A2). 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 that $q = s$, then there is no transition in the current period, and each player i receives stage utility $(1-\beta) w_i(s)$; in addition, if $\phi(s) = \phi(q) \neq q = s$, then there is a transition next period. In either case, the continuation payoffs are given by (A2).

Third, we show that there are no profitable deviations from σ^* at any stage. 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 stage constitutes a finite game with two possible outcomes. Lemma 1(a) then establishes that 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 stage, each player i compares continuation payoff $V_i(s, \phi(s))$ if the proposal is accepted and $V_i(s, s)$ if it is rejected. In all other voting stages, player i receives $V_i(s, P_k)$ if proposal P_k is accepted and $V_i(s, \phi(s))$ 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 $V_i(s, \phi(s)) = V_i(s, s)$ if $\phi(s) = s$). Therefore, there are no profitable deviations from σ^* given the continuation payoffs in (7). This argument establishes that the strategy profile σ^* is a best response to itself for any $s \in \mathcal{S}$ in the truncated game given the continuation payoffs in (7). Since we have already established that under σ^*, the continuation payoffs starting in state s are given by $V_i(s, q)$ in (A2), σ^* is a MPE of the entire game, which completes the proof of the first part of the Theorem.

(Part 2) We first prove that a MPE exists. We then show that it has the properties stated in part 2 of the Theorem.

Let us first construct a specific mapping ϕ satisfying Axioms 1–3. Take a sequence of states $\left\{\mu_1, ..., \mu_{|\mathcal{S}|}\right\}$ satisfying (7). Then, follow the procedure described in Theorem 1. First, we set $\phi(\mu_1) = \mu_1$. If for $k \geq 2$ we have $\mathcal{M}_k = \varnothing$, then $\phi(\mu_k) = \mu_k$; otherwise, let $Z_k \subset \mathcal{M}_k$ be defined as

$$Z_k = \left\{z \in \mathcal{M}_k : \forall s \in \mathcal{M}_k : s \not\succ_{\mu_k} z\right\}.$$

The set Z_k is nonempty by Assumption 2(b)*, and according to the procedure, any element of Z_k may be chosen as $\phi(\mu_k)$. Proceeding inductively, a specific mapping ϕ is obtained.

We construct an equilibrium in which continuation payoff of player i if the current state is s and proposal q is accepted, $V_i(s, q)$, is given by (A2) (in particular, if no alternative is accepted at a given period, each player i receives $V_i(s, s)$). Given these continuation payoffs, each period can be viewed as a finite (truncated) game with terminal payoffs given by $V_i(s, q)$. By backward induction, we can construct a SPE σ' of this truncated game as follows: let k^* be such that $\pi_s(k^*) = \phi(s)$ if such k^* exists; otherwise, let k^* be the first stage where $\pi_s(k^*) = i \in \mathcal{I}$ where $w_i(\phi(s)) > w_i(s)$ (such i exists, because ϕ satisfies Axiom 1). We require that $\phi(s)$ is proposed and accepted at stage k^*, and that no proposal is accepted at any stage $l < k^*$. Given the continuation payoffs in (A2), it is straightforward to verify that there are no profitable deviations from σ' and thus σ' is indeed a SPE. Since actions in σ' only depend on proposals and on the stage of this finite truncated game, we can choose σ' to be Markovian (the only requirement is to choose an SPE σ', where each player votes no when indifferent; clearly, such an SPE exists). Therefore, we have established the existence of an MPE.

We now establish the properties that any MPE satisfies. Take any set of sequences $\{\pi_s(\cdot)\}_{s\in\mathcal{S}}$ and any pure-strategy MPE σ. For any state s, the proposal q that is accepted along the equilibrium path is well-defined (let $q = s$ if all proposals are rejected) and let us denote it by $\chi(s) = q$. First, note that $\chi : \mathcal{S} \to \mathcal{S}$ has "no cycles," in the sense that if $\chi(s) \neq s$ then for any $n > 1$, $\chi^n(s) \neq s$ (where $\chi^2(s) \equiv \chi(\chi(s))$ etc.). This can be established by contradiction. Suppose there exists n such that $\chi^n(s) = s$, but $\chi(s) \neq s$. Denote by $J_s \subset \{1, \dots, K_s\}$ the set of voting stages in state s where a proposal P_k made along the 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). By the definition of the mapping χ, the first voting stage in J_s leads to $\chi(s)$. Now it suffices to consider two cases.

Case (i): all voting stages in J_s lead to cycles. Suppose this is the case and consider the last voting stage k'. Here each player knows that he will receive zero utility if $P_{k'}$ is accepted, and that he will receive $(1-\beta) w_s(i) > 0$ if $P_{k'}$ is accepted. Then Lemma 1(c) implies that $P_{k'}$ cannot be accepted in any MPE, thus yielding the desired contradiction.

Case (ii): not all voting stages in J_s lead to cycles. In this case, denote the voting stages that do not lead to cycles by $J_s' \subset J_s$. Consider the last voting stage k in J_s that precedes the first voting stage in J_s'. Accepting the proposal made at k', $P_{k'}$, leads to zero utility to each player, while rejecting it leads to a positive payoff. Therefore, proposal $P_{k'}$ cannot be accepted in any MPE, again yielding a contradiction and establishing the "no cycle" result.

This "no cycle" result in turn implies that $\chi^n(s) = \chi^{|\mathcal{S}|} - 1(s)$ for all $n \geq |\mathcal{S}| - 1$. Then, define $\psi(s) = \chi^{|\mathcal{S}|-1}(s)$, and

$$m(s) = \min\{n \in \mathbb{N} \cup \{0\} : \chi^n(s) = \psi(s)\}, \tag{A3}$$

(with $\chi^0(s) = s$). Evidently, $0 \leq m(s) \leq |\mathcal{S}| - 1$, and $m(s) = 0$ if and only if $\psi(s) = \chi(s) = s$. Moreover,

$$\psi(\psi(s)) = \chi(\psi(s)) = \psi(\chi(s)) = \psi(s) \tag{A4}$$

for any state s, as follows from the definition of mapping ψ). Finally, let us also define

$$\bar{V}_i(s, q) = \left\{ \begin{array}{cc} (1-\beta) w_i(s) & \text{if } s = q \\ 0 & \text{if } s \neq q \end{array} \right\} + \beta^{m(q)+1} w_i(\psi(q)). \tag{A5}$$

Clearly, $\bar{V}_i(s, a)$ gives the continuation payoff of player i if in state s alternative q is implemented, and subsequently equilibrium play (according to the MPE σ) follows. The rest of the proof involves showing that (1) $\psi(s)$ satisfies Axioms 1–3, and then (2) $\chi(s) = \psi(s)$ (this second statement is equivalent to showing that $\chi(s)$ is the dynamically stable state reached with zero or one transition, so that in the MPE σ, $s_t = \chi(s_0)$ for all $t \geq 1$). We start with an intermediate result. Then we prove that $\psi(s)$ satisfies Axioms 1–2. Then we prove that $\chi(s) = \psi(s)$. Finally, we prove that ψ satisfies Axiom 3. (This order makes the proof simpler).

Proof that if proposals P_{k_j} and P_{k_l} are proposed and accepted in state s, then $\psi\left(P_{k_j}\right) \sim \psi(P_{k_l})$. To establish this, for each state s consider again the set of voting stages J such that for each $k \in J$, the proposal P_k is accepted. Let $J = \left\{k_1, \ldots, k_{|J|}\right\}$, where $k_j < k_l$ for $j < l$ (we drop index s for convenience), and suppose that $J \neq \varnothing$ (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 $m(P_{k_1}) = m(s) - 1$ (which implies that $\psi(s) \neq s$). Since each P_{k_l} for $1 \leq l \leq |J|$ is accepted in this equilibrium, we must have, again by Lemma 1(c), that

$$\psi(P_{k_l}) \succeq_s \psi\left(P_{k_{l+1}}\right) \text{ for } 1 \leq l \leq |J| - 1$$

(in particular, only players who weakly prefer $\psi(P_{k_l})$ to $\psi\left(P_{k_{l+1}}\right)$ could vote for acceptance, since $w_i(\psi(P_{k_l})) < w_i\left(\psi\left(P_{k_{l+1}}\right)\right)$ is sufficient to imply $\bar{V}_i(s, P_{k_l}) < \bar{V}_i\left(s, P_{k_{l+1}}\right)$ in view of the fact that $\beta > \beta_0$).

In addition, we also have

$$\psi\left(P_{k_{|J|}}\right) \succeq_s \psi\left(P_{k_1}\right).$$

This can be seen as follows. If $P_{k_{|J|}}$ is accepted, each player i will receive $\bar{V}_i\left(s, P_{k_{|J|}}\right) = \beta^{m\left(P_{k_{|J|}}\right)+1} w_i\left(\psi\left(P_{k_{|J|}}\right)\right)$, while if it is rejected, each player will receive $\bar{V}_i(s,s) = (1-\beta) w_i(s) + \beta^{m(s)+1} w_i\left(\psi\left(P_{k_1}\right)\right)$. Now if $w_i\left(\psi\left(P_{k_{|J|}}\right)\right) < w_i\left(\psi\left(P_{k_1}\right)\right)$, then we would have $\beta^{m\left(P_{k_{|J|}}\right)+1} w_i\left(\psi\left(P_{k_{|J|}}\right)\right) < \beta^{m(s)+1} w_i\left(\psi\left(P_{k_1}\right)\right)$, and hence $\bar{V}_i\left(s, P_{k_{|J|}}\right) < \bar{V}_i(s,s)$. Since $P_{k_{|J|}}$ is accepted, the set of players for whom $w_i\left(\psi\left(P_{k_{|J|}}\right)\right) < w_i\left(\psi\left(P_{k_1}\right)\right)$ must be sufficiently small, and, more precisely, we must have

$$\left\{i \in \mathcal{I} : w_i\left(\psi\left(P_{k_{|J|}}\right)\right) \geq w_i\left(\psi\left(P_{k_1}\right)\right)\right\} \in \mathcal{W}_s.$$

This establishes that $\psi\left(P_{k_{|J|}}\right) \succeq_s \psi\left(P_{k_1}\right)$. Now, since Assumption 2(b)* holds, we have $\psi\left(P_{k_j}\right) \sim \psi\left(P_{k_l}\right)$ for all $1 \leq j < l \leq |J|$. In addition, we prove that $m\left(P_{k_l}\right) \leq m\left(P_{k_{l+1}}\right)$ 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.

Proof that ψ satisfies Axiom 1. Consider the 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

$$\bar{V}_i\left(s, P_{k_{|J|}}\right) = \beta^{m\left(P_{k_{|J|}}\right)+1} w_i\left(\psi\left(P_{k_{|J|}}\right)\right).$$

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

$$\bar{V}_i(s,s) = (1-\beta) w_i(s) + \beta^{m(s)+1} w_i(\psi(s)).$$

We have established, however, that $w_i\left(\psi\left(P_{k_{|J|}}\right)\right) = w_i(\psi(s))$ and that $m(s) = m\left(P_{k_1}\right) + 1 \leq P_{k_{|J|}} + 1$. Since $P_{k_{|J|}}$ is accepted in equilibrium, by Lemma 1(c) we must have that $\bar{V}_i\left(s, P_{k_{|J|}}\right) \geq \bar{V}_i(s,s)$ for a winning coalition of players in s, that is,

$$X_s = \left\{i \in \mathcal{I} : \bar{V}_i\left(s, P_{k_{|J|}}\right) \geq \bar{V}_i(s,s)\right\} \in \mathcal{W}_s.$$

Then for all $i \in X_s$, we must have

$$\beta^{m(s)} w_i(\psi(s)) \geq \beta^{m\left(P_{k_{|J|}}\right)+1} w_i(\psi(s)) \geq (1-\beta) w_i(s) + \beta^{m(s)+1} w_i(\psi(s)).$$

This implies that

$$\beta^{m(s)} w_i(\psi(s)) \geq w_i(s) \text{ for all } i \in X_s,$$

which, in view of the fact that $\beta > \beta_0$, implies that $w_i(\psi(s)) \geq w_i(s)$ for all $i \in X_s$. However, if for some $i \in X_s$, $w_i(\psi(s)) = w_i(s)$, then we would have $\beta^{m(s)} w_i(\psi(s)) < w_i(s)$, because $m(s) \geq 1$ and $\psi(s) \neq s$. Consequently, $w_i(\psi(s)) > w_i(s)$ for all $i \in X_s$, which implies that

$$\{i \in \mathcal{I} : w_i(\psi(s)) > w_i(s)\} \in \mathcal{W}_s,$$

thus establishing that

$$\psi(s) \succ_s s \text{ for any } s \in \mathcal{S} \text{ with } \psi(s) \neq s,$$

and therefore Axiom 1 holds.

Proof that ψ satisfies Axiom 2. This is straightforward in view of the fact that $\psi(\psi(s)) = \psi(s)$.

Proof that $\chi(s) = \psi(s)$. Let us 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, in (A3) $m(s) = 1$ whenever $\chi(s) \neq s$). Consider two cases.

Case (i): $\psi(s) = P_{k_j}$ for some $j : 1 \leq j \leq |J|$. In this case, $m\left(P_{k_j}\right) = 0$ since Axiom 2 is proven to hold. But we proved that $m(P_{k_l})$ is weakly increasing in l, therefore, $m(\chi(s)) = m(P_{k_1}) = 0$, and therefore $m(s) = 1$.

Case (ii): $\psi(s) = P_{k_j}$ does not hold for any j. This implies that $m(P_{k_1}) \geq 1$ and $\psi(s) \neq \chi(s)$. First observe 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. This can be established with the following argument: take any player i. If $P_k = \psi(s)$ is accepted, this player will receive $\bar{V}_i(s, \psi(s)) = \beta w_i(\psi(s))$, while if it is rejected, he will receive

$$\bar{V}_i(s, P_{k_l}) = \beta^{m\left(P_{k_l}\right)+1} w_i(\psi(s)) \leq \beta^2 w_i(\psi(s))$$

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

$$\bar{V}_i(s, s) = (1-\beta) w_i(s) + \beta^{m\left(P_{k_1}\right)+1} w_i(\psi(s)) \leq (1-\beta) w_i(s) + \beta^2 w_i(\psi(s))$$

if $k > k_{|J|}$. In the first case, all players prefer to have P_k accepted, while in the second case, each player with $w_i(\psi(s)) > w_i(s)$ will have $\beta w_i(\psi(s)) > w_i(s)$ since $\beta > \beta_0$, and therefore $\bar{V}_i(s, \psi(s)) > \bar{V}_i(s, s)$. Since such players form a winning coalition, we conclude that $P_k = \psi(s)$ will necessarily be accepted. Since we know that this $k \notin J$, it must be the case that proposal $\psi(s)$ is never considered on equilibrium path. By Assumption 4, it must be that each player become agenda-setter for some k. But then take any player i such that $w_i(\psi(s)) > w_i(s)$ and suppose that he is agenda-setter at stage k. This player's equilibrium proposal will give him utility $\bar{V}_i(s, P_{k_l})$ for some l if $k \leq k_{|J|}$ and $\bar{V}_i(s, s)$ if $k > k_{|J|}$. However, proposing $\psi(s)$ will give him a strictly higher utility $\bar{V}_i(s, \psi(s))$, as shown above. Therefore, player i has a profitable deviation. This contradiction shows that the case where $\psi(s) = P_{k_j}$ does not hold for any j is impossible, and thus finishes the proof that transition to state $\psi(s)$ takes place in one step, so that $\psi(s) = \chi(s)$.

Proof that ψ satisfies Axiom 3. Suppose, to obtain a contradiction, that Axiom 3 does not hold. This implies that there exists state $s, z \in \mathcal{S}$ such that $\psi(z) = z$, $z \succ_s s$ (which implies $z \neq s$), and $z \succ_s \psi(s)$ (which implies $\psi(z) \nsim \psi(s)$). As before, we can prove that if $P_k = z$ for some k, then proposal P_k must be accepted. In particular, accepting proposal z will lead to utility $\bar{V}_i(s, z) = \beta w_i(z)$ for any player i, while rejecting can lead to one of two possible payoffs. These possible payoffs are:

Case (i): $\bar{V}_i(s, P_{k_l}) \leq \beta w_i(\psi(s))$ for some l if $J \neq \varnothing$ and $k < k_{|J|}$;

Case (ii): $\bar{V}_i(s, s) = (1-\beta) w_i(s) + \beta^{m(s)+1} w_i(\psi(s)) \leq (1-\beta) w_i(s) + \beta w_i(\psi(s))$ if $J = \varnothing$ or $k > k_{|J|}$.

The fact that (by hypothesis) $z \succ_s \psi(s)$ implies that $\{i : w_i(z) > w_i(\psi(s))\} \in \mathcal{W}_s$ (that is, players that obtain higher stage payoff from z than from $\psi(s)$ form a winning coalition in s). In case (i), from (A5), we have

$$\bar{V}_i(s, z) = \beta w_i(z) > \beta w_i(\psi(s)) \geq \bar{V}_i(s, P_{k_l}).$$

In case (ii), because $\beta > \beta_0$ (recall (A1)), we have

$$\beta w_i(z) \geq (1-\beta) w_i(s) + \beta w_i(\psi(s)),$$

and therefore $\bar{V}_i(s, z) > \bar{V}_i(s, s)$.

We have therefore established that in both cases Lemma 1(c) applies and implies that proposal $P_k = z$ must be accepted. However, we have already shown that any proposals that are proposed and accepted is mapped (by ψ)to equivalent states. Hence, if z is ever proposed, we must have $\psi(z) \sim \psi(s)$. Since $\psi(z) \nsim \psi(s)$, it must be the case that z is not proposed.

Assumption 4 now implies that either each state is proposed or that each player becomes agenda-setter for some k. The former clearly cannot be the case, so suppose the latter applies. Consider player i for whom $w_i(z) > w_i(\psi(s))$ and suppose that he is the agenda-setter at some stage k. If he makes his equilibrium proposal, he receives either

$$\bar{V}_i(s, P_{k_l}) \leq \beta w_i(\psi(s)),$$

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

$$\bar{V}_i(s, s) = (1-\beta) w_i(s) + \beta^{m(s)+1} w_i(\psi(s)) \leq (1-\beta) w_i(s) + \beta w_i(\psi(s)),$$

depending on $k \leq k_{|J|}$ or $k > k_{|J|}$. Instead, if he proposes $P_k = z$, he will receive

$$\bar{V}_i(s,z) = \beta w_i(z) > \max\{\beta w_i(\psi(s)), (1-\beta) w_i(s) + \beta w_i(\psi(s))\},$$

where the inequality follows from $w_i(z) > w_i(\psi(s))$ and $\beta > \beta_0$. This implies that player i has a profitable deviation, yielding a contradiction. This establishes that ψ satisfies Axiom 3, and thus completes the proof of part 2 of the Theorem.

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

7.3 Proof of Theorem 3

The first step is again a key lemma, which is of potential independent interest. This lemma characterizes properties of quasi-median voters under more general political institutions (parallel to the properties of median voters in majoritarian elections). For this lemma, recall that M_s denotes the set of quasi-median voters in state s.

Lemma 2 *Given $\mathcal{I} \subset \mathbb{R}$, $\mathcal{S} \subset \mathbb{R}$, payoff functions $\{w_i(s)\}_{i\in\mathcal{I},s\in\mathcal{S}}$, and winning coalition $\{\mathcal{W}_s\}_{s\in\mathcal{S}}$ satisfying Assumption 1, the following are true.*

1. *For each s, the set M_s is nonempty.*

2. *If the single-crossing property (SC) in Definition 5 holds, then for any states $x, y, z \in \mathcal{S}$,*

$$x \succ_z y \Leftrightarrow w_i(x) > w_i(y) \text{ for all } i \in M_z, \text{ and}$$

$$x \succeq_z y \Leftrightarrow w_i(x) \geq w_i(y) \text{ for all } i \in M_z.$$

3. *If monotonic median voter condition in Definition 8 holds, then there exists a sequence $\{m_s\}_{s\in\mathcal{S}}$ of players such that $m_s \in M_s$ for all $s \in \mathcal{S}$ and whenever states $x, y \in \mathcal{S}$ satisfy $x < y$, $m_x \leq m_y$.*

Proof. **(Part 1)** Let b be such that $B = \{j \in \mathcal{I} : -\infty < j \leq b\} \in \mathcal{W}_s$ and $\{j \in \mathcal{I} : -\infty < j < b\} \notin \mathcal{W}_s$. Intuitively, such B is the "leftmost" winning coalition. Similarly, let a be such that $A = \{j \in \mathcal{I} : a \leq j < \infty\} \in \mathcal{W}_s$ and $\{j \in \mathcal{I} : a < j < \infty\} \notin \mathcal{W}_s$, so that A is the "rightmost" winning coalition. Assumption 1 implies that $Z = A \cap B \neq \varnothing$. Since all quasi-median voters must be both in A and B, we also have $M_s \subset Z$. Next, we show that $Z \subset M_s$ is also true. To obtain a contradiction, assume the opposite. Then for some "connected" coalition $X = \{j \in \mathcal{I} : x \leq j \leq y\} \in \mathcal{W}_s$ the inclusion $Z \subset X$ does not hold. Then, evidently, either the lowest or the highest quasi-median voter is not in X. Suppose, without loss of generality, the latter is the case. Since X is winning, coalition $Y = \{j \in \mathcal{I} : -\infty < j \leq y\}$

(where y is the highest player in X) is winning, and therefore $Z \subset Y$. But this implies that the highest quasi-median voter is neither in X nor in Y, which is impossible and thus yields a contradiction. This proves that $M_s = Z \neq \varnothing$.

(Part 2) Consider the case $x \geq y$ (the case $x < y$ is treated similarly). Suppose $x \succ_z y$. Then $\{i \in \mathcal{I} : w_x(i) > w_y(i)\} \in \mathcal{W}_z$ (is winning in z). But by SC, this coalition is connected, and therefore includes all players from M_z. Conversely, suppose that $w_i(x) > w_i(y)$ for all $i \in M_z$. Now SC implies that the same inequality holds for player j whenever $j \geq i \in M_z$. Part 1 of the Lemma implies that $\{j \in \mathcal{I} : \exists i \in M_z \text{ such that } j \geq i\} \in \mathcal{W}_z$. This establishes that $w_i(x) > w_i(y)$ for all $i \in M_z$ implies $x \succ_z y$, and completes the proof for this case. The proof of the results for the $\succeq$ relation is analogous.

(Part 3) By part 1 of this Lemma, the set M_s is nonempty for each $s \in \mathcal{S}$. Let

$$m_s = \max_{x \in \mathcal{S} : x \leq s} \min_{m \in M_x} m. \tag{A6}$$

Evidently, if $x < y$, then $m_x \leq m_y$. Moreover, $m_s \in M_s$. To prove this last statement, assume the opposite; then $m_s = \min_{m \in M_x}$ for some $x < s$. Since we assumed $m_s \notin M_s$, then either $m_s \in M_x$ is less than all elements in M_s or greater than all elements in M_s. In the first case, $m_s < \min_{m \in M_s} m$, which violates the definition of m_s in (A6). In the second case, we find that M_s lies to the left of M_x, violating the monotonic median voter property. This contradiction proves that $m_s \in M_s$ for all $s \in \mathcal{S}$. Since the sequence (A6) is increasing, part 3 follows. ■

Proof of Theorem 3 (Part 1) We start with Assumption 2(a). Suppose that it does not hold, and there is a cycle $s_1, \ldots, s_l$ such that $s_{k+1} \succ_{s_k} s_k$ for $1 \leq k \leq l-1$ and $s_1 \succ_{s_l} s_l$. Take a monotonic sequence of median voters $\{m_s\}_{s \in \mathcal{S}}$. Recall that m_s is part of any connected winning coalition in s, therefore, if for some x and z, $x \succ_z z$, then $w_x(m_z) > w_z(m_z)$. Now for each $s \in \mathcal{S}$ consider an alternative set of winning coalitions where m_s is the dictator, i.e., $\mathcal{W}'_s = \{X \in \mathcal{C} : m_s \in X\}$. Denoting the induced relation between states by $\succ'$, we have that if $x \succ_z z$, then $x \succ'_z z$. Consequently, if there was a cycle $s_1, \ldots, s_l$ such that $s_{k+1} \succ_{s_k} s_k$ for $1 \leq k \leq l-1$ and $s_1 \succ_{s_l} s_l$, then we have $s_{k+1} \succ'_{s_k} s_k$ for $1 \leq k \leq l-1$ and $s_1 \succ'_{s_l} s_l$; therefore, a cycle for $\succ'$ exists. Now take the shortest cycle for $\succ'$ (this need not be a cycle for $\succ$). Without loss of generality, suppose that s_2 is the lowest state (so $s_2 \leq s_1$ and $s_2 \leq s_3$; then $m_{s_2} \leq m_{s_1}$ and $m_{s_2} \leq m_{s_3}$. Since $s_3 \succ'_{s_2} s_2$ and $s_2 \succ'_{s_2} s_1$, we have $w_{m_{s_2}}(s_3) > w_{m_{s_2}}(s_2)$ and $w_{m_{s_1}}(s_2) > w_{m_{s_1}}(s_1)$. But $s_2 \leq s_3$ and $m_{s_2} \leq m_{s_1}$, hence, $w_{m_{s_2}}(s_3) - w_{m_{s_2}}(s_2) > 0$ implies $w_{m_{s_1}}(s_3) - w_{m_{s_1}}(s_2) > 0$, i.e., $w_{m_{s_1}}(s_3) > w_{m_{s_1}}(s_2)$. Combining this with $w_{m_{s_1}}(s_2) > w_{m_{s_1}}(s_1)$, we conclude that $w_{m_{s_1}}(s_3) > w_{m_{s_1}}(s_1)$. But then $s_3 \succ'_{s_1} s_1$, since m_{s_1} is the dictator in s_1. This implies that the link s_2 may be skipped in the cycle, which contradicts the assumption that the cycle $\{s_k\}_{k=1}^{l}$ was the shortest one. This contradiction establishes that Assumption 2(a) holds.

To show that Assumption 2(b) holds, take any $s \in \mathcal{S}$ and some $m_s \in M_s$. From Lemma 2 it follows that if for some x, y we have $x \succ_s y$, then $w_{m_s}(x) > w_{m_s}(y)$. Suppose, to obtain a contradiction, that there is a cycle $s_1, \ldots, s_l$ such that $s_{k+1} \succ_s s_k$ for $1 \leq k \leq l-1$ and $s_1 \succ_s s_l$. Without loss of generality, we may assume that state s_l maximizes the payoff of m_s among states $s_1, \ldots, s_l$. This means that $w_{m_s}(s_l) \geq w_{m_s}(s_1)$, which implies $s_1 \not\succ_s s_l$ and thus contradicts the existence of a cycle.This shows that Assumption 2(b) holds and completes the proof of part 1.

(Part 2) Note that if preferences of player i are single-peaked, then his preferences' restricted to any subset $\mathcal{Q}$ of $\mathcal{S}$ are also single-peaked. For any nonempty subset $\mathcal{Q} \subset \mathcal{S}$ and $i \in \mathcal{I}$, let

$$b_i(\mathcal{Q}) \in \arg\max_{s \in \mathcal{Q}} w_i(s) \tag{A7}$$

(in case this maximum is achieved at multiple states, we pick any of these).

We start with Assumption 2(a). Suppose there is a cycle $s_1, \ldots, s_l$ such that $s_{k+1} \succ_{s_k} s_k$ for $1 \leq k \leq l-1$ and $s_1 \succ_{s_l} s_l$. Let us re-enumerate players in $\mathcal{I}$ as $i_1, \ldots, i_{|\mathcal{I}|}$ so that $b_{i_k}(\{s_1, \ldots, s_l\})$ is nondecreasing in k. It is straightforward to use the assumption that any two winning coalitions, even for different states, intersect, and prove that for this order there exists a quasi-median voter i_m such that $i_m \in X$ for any $X \in \mathcal{C}$ that satisfies $X \in \mathcal{W}_{s_k}$ for some $1 \leq k \leq l$ and $X = \{i_j \in \mathcal{I} : i_p \leq i_j \leq i_q\}$ for some $i_p, i_q \in \mathcal{I}$. Let $z = b_{i_m}(\{s_1, \ldots, s_l\})$ be the favorite state of quasi-median voter i_m. Then there exists s_k such that $s_k \succ_z z$. Without loss of generality, assume $s_k > z$. Because preferences are single-peaked, all players i_p with $p \leq m$, including i_m, (weakly) prefer z to s_k. By the choice of i_m, players who strictly prefer s_k to z do not form a winning coalition, which is a contradiction proving that Assumption 2(a) holds in this case.

Now, suppose that Assumption 2(b) is violated, i.e., there exist $s \in \mathcal{S}$ and a cycle $s_1, \ldots, s_l$ such that $s'_{k+1} \succ_s s_k$ for $1 \leq k \leq l-1$ and $s_1 \succ_s s_l$. Again, we re-enumerate players in $\mathcal{I}$, so that $b_{i_k}(\{s_1, \ldots, s_l\})$ is nondecreasing for this new cycle $s_1, \ldots, s_l$ and choose i_m such that $i_m \in X$ for any $X \in \mathcal{W}_s$ such that $X = \{i_j \in \mathcal{I} : i_p \leq i_j \leq i_q\}$ for some $i_p, i_q \in \mathcal{I}$. Now take $z = b_{i_m}(\{s_1, \ldots, s_l\})$; then there exists s_k such that $s_k \succ_s z$. Without loss of generality, assume $s_k > z$. But then $s_k \succ_s z$ is impossible, since all players i_p with $p \leq m$ (weakly) prefer z to x. This proves that Assumption 2(a,b) holds.

(Part 3) The case of part 1: The first part of Assumption 2(b)* is shown similarly to Assumption 2(b), making use of Assumption 5 to rule out indifferences between x and y when $x \succeq_z y$. Finally, if $x, y \in \mathcal{S}$ are such that $x \succ_s s$ and $y \succ_s x$, then for any $i \in M_s$ we have $w_i(y) > w_i(x) > w_i(s)$, which, in turn, implies $y \succ_s s$.

The case of part 2: The first part of Assumption 2(b)* is proved with an argument analogous to part 2, but also making use of Assumption 5. To prove the last part, take states x, y, s such that $x \succ_s s$ and $y \succ_s x$; this implies that these three states are different. Take the median voter $m_s \in M_s$. We must have $b_{m_s}(\{x, y, s\}) = y$ (with b defined by (A7)). To see this note that if $b_{m_s}(\{x, y, s\}) = s$, then $x \succ_s s$ would be impossible, and if $b_{m_s}(\{x, y, s\}) = s$, then $y \succ_s x$ would not hold. Now consider two cases. First, suppose that either $x, s > y$ or $x, s < y$. Without loss of generality assume $x, s > y$, in which case $x \succ_s s$ implies $y < x < s$. Now, we have that $w_i(y) > w_i(x)$ if and only if $b_i(\{x, y, s\}) = y$, so such players form a winning coalition in s. This implies $y \succ_s s$. Second, suppose that either $x < y < s$ or $s < y < x$. Without loss of generality assume the former. Now, if a player prefers x to s, he must also prefer y to s; given that $x \succ_s s$, we must therefore have $y \succ_s s$. In both cases, we have $y \succ_s s$, which completes the proof. ■

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, New York.

Acemoglu, Daron, and James Robinson (2006b) "Economic Backwardness in Political Perspective" *American Political Science Review*, 100(1), 115-131.

Alesina, Alberto, Ignazio Angeloni, and Federico Etro (2005) "International Unions," *American Economic Review*, 95(3), 602-615.

Austen-Smith, David and Jeffrey Banks (1988) "Elections, Correlations and Legislative Outcomes," *American Political Science Review* 82 405-422.

Barbera, Salvador and Matthew Jackson (2004) "Choosing How to Choose: Self-Stable Majority Rules and Constitutions," *Quarterly Journal of Economics*, 119(3), 1011-1048.

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 Bernardo Moreno (2008) "Top Monotonicity: A Common Root for Single

Peakedness, Single Crossing, and the Median Voter Result," mimeo.

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.

Diermeier, Daniel and Pohan Fong (2008) "Endogenous Limits on Proposal Power," mimeo.

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.

Hirshleifer, David, Michelle Boldrin, and David K. Levine (2008) "The Slippery Slope of Concessions," mimeo.

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.

Konishi, Hideo, and Debraj Ray (2003), "Coalition Formation as a Dynamic Process," *Journal of Economic Theory*, 110: 1-41.

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), 705-763.

Messner, Matthias and Mattias 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.

von Neumann, John and Oskar Morgenstern (1944) *Theory of Games and Economic Behavior,* Princeton University Press.

Appendix B: Limited Transitions (Not for Publication)

7.4 Modeling Limited Transitions

In the main body of the paper we assumed that any transition (from any state to any other state) is possible. In certain applications, not all transitions across states may be 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), also discussed in subsection 6.4. 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. Here in Appendix B we show that our analysis, after proper reformulation of the Assumptions and the Axioms of Section 3, is applicable to the case where 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 (similarly, $\mathcal{Q}_1 \rightsquigarrow \mathcal{Q}_2$). The analysis in the main body of the paper thus corresponds to the special case where $\mathcal{S} \rightsquigarrow \mathcal{S}$ for any $x \in \mathcal{S}$. We adopt the following natural assumption on the transition relation.

Assumption 6 ***(Feasible Transitions)*** *Relation $\rightsquigarrow$ satisfies the following properties:*

(a) (reflexivity) $\forall x \in \mathcal{S} : x \rightsquigarrow x$*;*

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

Part (b) Assumption 6 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.[16] 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).[17]

[16] See Chwe (1994) for another model where different transitions require different winning coalitions.

[17] One set of economically interesting cases in which Assumption 6 fails 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 6 is natural here. In particular,

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).

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

(a) For any sequence of states $s_1, \ldots, s_k$ *in* $\mathcal{S}$ *with* $s_j \rightsquigarrow s_{j+1}$ *for all* $1 \leq j \leq k-1$ *and* $s_k \rightsquigarrow s_1$,

$$s_{j+1} \succ_{s_j} s_j \text{ for all } 1 \leq j \leq k-1 \Longrightarrow s_1 \not\succ_{s_k} s_k.$$

(b) For any sequence of states $s, s_1, \ldots, s_k$ *in* $\mathcal{S}$ *with* $s \rightsquigarrow s_j$ *for all* $1 \leq j \leq k$, $s_k \rightsquigarrow s_1$, *and* $s_j \succ_s s$ *for all* $1 \leq j \leq k$,

$$s_{j+1} \succ_s s_j \text{ for all } 1 \leq j \leq k-1 \Longrightarrow s_1 \not\succ_s s_k.$$

(b) For any sequence of states* $s, s_1, \ldots, s_k$ *in* $\mathcal{S}$ *with* $s \rightsquigarrow s_j$ *for all* $1 \leq j \leq k$, $s_j \not\sim s_l$ *for some* $1 \leq j < l \leq k$, *and* $s_j \succ_s s$ *for all* $1 \leq j \leq k$,

$$s_{j+1} \succeq_s s_j \text{ for all } 1 \leq j \leq k-1 \Longrightarrow s_1 \not\succeq_s s_k.$$

Moreover, if for $x, y, s \in \mathcal{S}$ *we have* $s \rightsquigarrow x$, $s \rightsquigarrow y$, $x \succ_s s$ *and* $y \not\succ_s s$, *then* $y \not\succ_s x$.

Assumption $3'$ ***(Comparability with Limited Transitions)*** *For* $x, y, s \in \mathcal{S}$ *such that* $s \rightsquigarrow x$, $s \rightsquigarrow y$, $x \succ_s s$, $y \succ_s s$, *and* $x \not\sim y$, *either* $y \succ_s x$ *or* $x \succ_s y$.

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 $2'$ ***(Stability)*** *If* $x, y \in \mathcal{S}$ *are such that* $y = \phi(x)$, *then* $y = \phi(y)$.

Axiom $3'$ ***(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* $y \neq \phi(x)$.

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

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).

Theorem 4 ***(Dynamically Stable States with Limited Transitions)*** *Suppose that binary relation $\rightsquigarrow$ satisfies Assumption 6, and that Assumptions 1 and $2'$ hold. Then:*

1. *There exists mapping ϕ satisfying Axioms $1'$–$3'$.*

2. *Any mapping ϕ that satisfies Axioms $1'$–$3'$ can be recursively constructed as follows. Construct sequence $\left\{\mu_1, ..., \mu_{|\mathcal{S}|}\right\}$ with the property that for any $1 \leq j < l \leq |\mathcal{S}|$, either $\mu_j \not\rightsquigarrow \mu_l$ or $\mu_l \not\succ_{\mu_j} \mu_j$.Let $\phi(\mu_1) = \mu_1$. For each $k = 2, \ldots, |\mathcal{S}|$, let*

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

and define

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

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) $\mu : \{1, \ldots, |\mathcal{S}|\} \leftrightarrow \mathcal{S}$ such that for any $1 \leq k < |\mathcal{S}|$ we have that

$$\text{if } 1 \leq j < l \leq |\mathcal{S}|\text{, then } \mu_j \not\rightsquigarrow \mu_l \text{ or } \mu_l \not\succ_{\mu_j} \mu_j. \tag{B1}$$

To construct mapping μ, we introduce a binary relation $\leftrightsquigarrow$ defined as

$$x \leftrightsquigarrow y \text{ if and only if } x \rightsquigarrow y \text{ and } y \rightsquigarrow x.$$

Assumption 6 guarantees that $\leftrightsquigarrow$ is an equivalence relation, inducing equivalence classes $\{\mathcal{E}_x\}_{x \in \mathcal{S}}$ defined as

$$\mathcal{E}_x = \{y \in \mathcal{S} : x \leftrightsquigarrow y\}$$

to be such that $\mathcal{E}_x$ and $\mathcal{E}_y$ either coincide or do not intersect. 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 $\mathcal{E}^1, \ldots, \mathcal{E}^l$ 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 $\mathcal{E}^1, \ldots, \mathcal{E}^m$ (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 $\mu_1^k, \ldots, \mu_{|\mathcal{E}^k|}^k$ so that $\mu_l^k \not\succ_{\mu_j^k} \mu_j^k$ 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 $\mu^1 \equiv \left(\mu_1^1, \ldots, \mu_{|\mathcal{E}^1|}^1\right)$, 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 (B1). The rest of the proof closely follows the one of Theorem 1 and is omitted. ■

Similarly, an equivalent of Theorem 2 again applies.

Theorem 5 ***(Noncooperative Foundations of Dynamically Stable States with Limited Transitions)*** *Suppose that binary relation $\rightsquigarrow$ satisfies Assumption 6, that Assumptions 1, 2′(a),(b) and 4 hold. Then there exists $\beta_0 \in [0, 1)$ such if the discount factor $\beta > \beta_0$, then:*

1. *For any mapping $\phi : \mathcal{S} \to \mathcal{S}$ satisfying Axioms 1′–3′ there exists a set of protocols $\{\pi_s\}_{s \in \mathcal{S}}$ and a pure-strategy 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 protocols $\{\pi_s\}_{s \in \mathcal{S}}$ there exists a pure-strategy MPE. Any such MPE σ has the property that for any initial state $s_0 \in \mathcal{S}$, $s_t = s^\infty$ for all $t \geq 1$. Moreover, there exists mapping $\phi : \mathcal{S} \to \mathcal{S}$ satisfying 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 protocols $\{\pi_s\}_{s \in \mathcal{S}}$, any pure-strategy MPE 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.

7.5 Generalization of Proposition 4

We now show how Proposition 4 can be generalized by allowing only certain types of transitions. This generalization makes the political elimination model of Acemoglu, Egorov, and Sonin (2008) (and its extension to infinite horizon) also a special case of the analysis here.

Proposition 5 *Consider the environment in Acemoglu, Egorov, and Sonin (2008). Then:*

1. *Assumptions 1, 2′, 3′, and 6 are satisfied (provided that $X \rightsquigarrow Y$ is feasible if and only if $Y \subset X$).*

2. *There exists an arbitrarily small perturbation of payoffs such that Assumption 2′(b)* holds.*

3. *If only eliminations are possible, 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).*

4. *In the case where any transition is feasible (as in Proposition 4 in the text), there exists a unique outcome mapping ϕ that satisfies Axioms 1–3. This mapping is potentially different from ϕ_{elim}.*

Proof. **(Part 1)** Assumption 1 is satisfied by Proposition 4 (part 1), since it has not changed. Assumptions $2'$ and $3'$ are in fact weaker than corresponding Assumptions 2 and 3, because in the latter two, all transitions are allowed, which makes the set of potential cycles larger. 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, so Assumption 6 is satisfied.

(Part 2) The proof of this result is very similar to the proof of part 2 of Proposition 4 and is omitted.

(Part 3) Existence and uniqueness of a mapping ϕ_{elim} that satisfies Axioms $1'$–$3'$ follows from part 1, since Theorem 1 is applicable. Similarly, in Acemoglu, Egorov, and Sonin (2008) it is shown that under these assumptions, there exists a unique outcome correspondence ϕ_{aes} that satisfy Axiom 1–4 of that paper This correspondence ϕ_{aes} is single-valued due to genericity assumption, so below, we will treat ϕ_{aes} as a mapping, not a correspondence. To prove that $\phi_{elim} = \phi$ it suffices to show that mapping ϕ_{elim} satisfies Axiom 1–4 of that paper. For Axiom 1 (Inclusion): Take any X; Axiom $1'$ implies that either $\phi_{elim}(X) = X$ or $X \rightsquigarrow \phi_{elim}(X)$. In both cases, $\phi_{elim}(X) \subset X$. For Axiom 2 (Power): Again, Axiom $1'$ implies that either $\phi_{elim}(X) = X$ or $X \rightsquigarrow \phi_{elim}(X)$ and $\phi_{elim}(X) \succ_X X$. In both cases, $\phi_{elim}(X) \subset X$ and $\gamma_{\phi_{elim}(X)} > \alpha\gamma_X$, so $\phi_{elim}(X) \in \mathcal{W}_X$ in the notation of that paper. Axiom 3 (Self-Enforcement): Take any X and let $Y = \phi_{elim}(X)$. Axiom $2'$ implies $\phi_{elim}(Y) = \phi_{elim}(\phi_{elim}(X)) = \phi_{elim}(X) = Y$. Finally, Axiom 4 (Rationality): Take any X, $Y = \phi_{elim}(X)$, and suppose that $Z \in \mathcal{W}_X$ (meaning that $Z \subset X$ and $\gamma_Z > \alpha\gamma_X$) and $Z = \phi_{elim}(Z)$. If $\gamma_Y < \gamma_Z$, then we have $X \rightsquigarrow Y$, $Y \succ_X X$, $Y = \phi_{elim}(Y)$ by Axiom $2'$, and $Y \succ_X Z$. Axiom $3'$ then implies that $Z \neq \phi_{elim}(X)$. Conversely, if $\gamma_Z \leq \gamma_Y$, then $\gamma_Z = \gamma_Y$, for $\gamma_Z < \gamma_Y$ would imply that $X \rightsquigarrow Z$, $Z \succ_X X$, $Z = \phi_{elim}(Z)$, and $Z \succ_X Y$, in which case we would get a contradiction with $Y = \phi_{elim}(X)$, as Axiom $3'$ would imply $Y \neq \phi_{elim}(Y)$. But $\gamma_Z = \gamma_Y$, together with genericity, implies $Z = Y$, and hence $Z = \phi_{elim}(X)$. This proves that Axioms 1–4 of Acemoglu, Egorov, and Sonin (2008) are satisfied, hence, $\phi_{elim} = \phi_{aes}$.

(Part 4) Existence and uniqueness follow from Proposition 4 (part 1). To show that the two mappings, ϕ_{elim} and ϕ may be different, consider four individuals, A, B, C, Z with powers $\gamma_A = 3$, $\gamma_B = 4$, $\gamma_C = 5$, $\gamma_D = 4.5$. It is straightforward to verify that $\phi_{elim}(\{A, B, C\}) = \{A, B, C\}$, but $\phi(\{A, B, C\}) = \{A, B, Z\}$. Moreover, it is possible that for some coalition X, $\phi_{elim}(X) \subset X$ and $\phi(X) \subset X$, but $\phi(X) \neq \phi_{elim}(X)$: this would be the case for six players A, B, C, D, E, F with powers $100, 101, 103, 107, 115, 131$, respectively (here, $\phi_{elim}(\{A, B, C, D, E, F\}) = \{A, B, F\}$, but $\phi(\{A, B, C, D, E, F\}) = \{D, E, F\}$). In fact, $\{A, B, F\}$ is the winning coalition in $\{A, B, C, D, E, F\}$ with the least power; it also happens to be ϕ_{elim}-stable. However, it is not ϕ-stable ($\phi(\{A, B, F\}) = \{A, B, C\}$), and in this case $\{D, E, F\}$ is the ϕ-stable winning coalition with the least total power (in $\{A, B, C, D, E, F\}$). ■

Appendix C: Examples, Proofs From Section 6 and Additional Results (Not for Publication)

Definition of MPE

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 $a_i = \left(a_i^1, a_i^2, \ldots\right)$ for $i = 1, \ldots, n$, with $a_i^t \in A_i^t$ and $a_i \in A_i = \prod_{t=1}^{\infty} A_i^t$. Let $h^t = \left(a^1, \ldots, a^t\right)$ be the history of play up to stage t (not including stage t), where $a^s = (a_1^s, \ldots, a_n^s)$, so h^0 is the history at the beginning of the game, and let H^t be the set of histories h^t for $t : 0 \leq t \leq T - 1$.

We denote the set of all potential histories up to date t by

$$H_t = \bigcup_{s=0}^{t} H^s.$$

Let t-continuation action profiles be $a_{i,t} = \left(a_i^t, a_i^{t+1}, \ldots\right)$ for $i = 1, \ldots, n$, with the set of continuation action profiles for player i denoted by $A_{i.t}$. Symmetrically, define t-truncated action profiles as $a_{i,-t} = \left(a_i^1, a_i^2, \ldots, a_i^{t-1}\right)$ for $i = 1, \ldots, n$, with the set of t-truncated action profiles for player i denoted by $A_{i,-t}$. We also use the standard notation a_i and a_{-i} to denote the action profiles for player i and the action profiles of all other players (similarly, A_i and A_{-i}). The payoff functions for the players depend only on actions, i.e., player i's payoff is given by $u_i\left(a^1, \ldots, a^n\right)$. A pure strategy for player i is

$$\sigma_i : H_\infty \to A_i.$$

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

$$\sigma_{i,t} : H_\infty \setminus H_{t-2} \to A_{i,t},$$

where $H_\infty \setminus H_{t-2}$ is the set of histories starting at time t.

We then have:

Definition 9 ***(Markovian Strategies)*** *A continuation strategy $\sigma_{i,t}$ is* ***Markovian*** *if*

$$\sigma_{i,t}\left(h_{t-1}\right) = \sigma_{i,t}\left(\tilde{h}_{\tau-1}\right)$$

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 $a_{-i,t} \in A_{-i,t}$,

$$u_i\left(a_{i,t}, a_{-i,t} \mid h_{t-1}\right) \geq u_i\left(\tilde{a}_{i,\tau}, a_{-i,t} \mid h_{\tau-1}\right)$$

implies

$$u_i\left(a_{i,t}, a_{-i,t} \mid \tilde{h}_{t-1}\right) \geq u_i\left(\tilde{a}_{i,\tau}, a_{-i,t} \mid \tilde{h}_{\tau-1}\right).$$

Markov perfect equilibria in pure strategies are defined formally as follows:

Definition 10 ***(MPE)*** *A pure strategy profile $\hat{\sigma} = (\hat{\sigma}_1, ..., \hat{\sigma}_n)$ is* ***Markov perfect equilibrium (MPE)*** *(in pure strategies) if each strategy $\hat{\sigma}_i$ is Markovian and*

$$u_i(\hat{\sigma}_i, \hat{\sigma}_{-i}) \geq u_i(\hat{\sigma}_i, \hat{\sigma}_{-i}) \text{ for all } \sigma_i \in \Sigma_i \text{ and for all } i = 1, \ldots, n.$$

Examples

Example 3 ***(Nonexistence without Transaction Costs)*** In this example, we show that a MPE in pure strategies may fail to exist if we assume away the transaction cost. There are 8 states $\mathcal{S} = \{A, B, C, D, E, F, G, H\}$ and 7 players. The set of winning coalitions are: $\mathcal{W}_A = \{X \in \mathcal{C} : |\{1, 2, 3\} \cap X| \geq 2\}$ (i.e., majority voting between $1, 2, 3$), $\mathcal{W}_B = [4]$, $\mathcal{W}_D = [5]$, $\mathcal{W}_F = [6]$, $\mathcal{W}_C = \mathcal{W}_E = \mathcal{W}_G = \mathcal{W}_H = [7]$ (here, $[i]$ denotes the set of winning coalitions where i is the dictator, so $[i] = \{X \in \mathcal{C} : i \in X\}$). The payoffs are as follows: $w_1(\cdot) = (0, 30, 0, 0, 20, 0, 0, 1)$, $w_2(\cdot) = (0, 0, 0, 30, 0, 0, 20, 1)$, $w_3(\cdot) = (0, 0, 20, 0, 0, 30, 0, 1)$, $w_4(\cdot) = (0, 0, 1, 0, 0, 0, 0, 0)$, $w_5(\cdot) = (0, 0, 0, 0, 1, 0, 0, 0)$, $w_6(\cdot) = (0, 0, 0, 0, 0, 0, 1, 0)$, $w_7(\cdot) = (0, 0, 0, 0, 0, 0, 0, 1)$. It is straightforward to show that Assumptions 1, 2 and 2b* are satisfied (it is helpful to notice that the only state s that satisfies $s \succ_A A$ is $s = H$).

Evidently, state H is stable (dictator 7 will never deviate), and similarly any of the states E, F, G will immediately lead to H. It is also evident that B will immediately lead to C, because C is the only state where dictator 4 receives a positive utility; similarly, D immediately leads to E and F immediately leads to G. Let us prove that no move from state A can form a pure-strategy equilibrium. First, it is impossible to stay in A: players $1, 2, 3$ would be better off moving to H. Moving to H immediately is not possible in an equilibrium either: Then players 1 and 3 would rather deviate and move to B, which would then lead to C and only then to H, since the average payoff of this path would be higher for each of these players (recall that the discount factor is close to 1).

Let us consider possible moves to B and C (the moves to D, E, F, G are considered similarly). If the state were to change to C, then players 1 and 2 would rather deviate and move to D (and then to E, followed by H). Finally, if the state were to change to B, then 2 and 3 could deviate to F, so as to follow the path to G and H after that; this is better for these players than moving to B, followed by C and H. So, without imposing a transaction cost it is possible that a pure-strategy equilibrium does not exist.

Example 4 ***(Cycles without Transaction Costs)*** In this example, we show that in the absence of transaction cost, an equilibrium may involve a cycle even though Assumptions 1, 2 and 2b* hold. There are 6 players, $\mathcal{I} = \{1,2,3,4,5,6\}$, and 3 states, $\mathcal{S} = \{A,B,C\}$. Players' preferences are given by $w_1(A,B,C) = (5,10,4)$, $w_2(A,B,C) = (5,4,10)$, $w_3(A,B,C) = (4,5,10)$, $w_4(A,B,C) = (10,5,4)$, $w_5(A,B,C) = (10,4,5)$, $w_3(A,B,C) = (4,10,5)$, and winning coalitions are defined by $\mathcal{W}_A = \{X \in \mathcal{C} : 1,2 \in X\}$, $\mathcal{W}_B = \{X \in \mathcal{C} : 3,4 \in X\}$, $\mathcal{W}_C = \{X \in \mathcal{C} : 5,6 \in X\}$. Then one can see that there is an equilibrium which involves moving from state A to state B, from B to C, and from C to A. To see this, because of the symmetry it suffices to see that the players will not deviate if the current state is A. The alternatives are to stay in A or move to C. But staying in A hurts both player 1 and player 2 (for player 2 who dislikes state B this is true because it postpones the move to C, the state that he likes best, while for player 1 this is evident). At the same time, moving to C hurts player 1, because state C is the worst of the three states for him not only in terms of instantaneous payoff, but also in terms of discounted present value (if the cycle continues, as it should due to the one-stage deviation principle). So, this cycle constitutes a (Markov Perfect) equilibrium.

It is also easy to see that in this example, Assumptions 1, 2 and 2b* are satisfied: in fact, there are no two states $s, s_0 \in \{A,B,C\}$ such that $s \succ_{s_0} s_0$. Finally, notice that the aforementioned cycle is not the only equilibrium. In particular, the cycle in the opposite direction may also arise in an equilibrium (this holds because of symmetry), and situation where all three states are stable is also possible (indeed, if B and C are stable, then players 1 will always block transition from A to C whereas player 2 will always block transition from A to B).

Example 5 ***(Nonexistence without Assumption 2(a))*** There are 3 players, $\mathcal{I} = \{1,2,3\}$, and 3 states, $\mathcal{S} = \{A,B,C\}$. Players' preferences satisfy $w_1(A) > w_1(B) > w_1(C)$, $w_2(B) > w_2(C) > w_2(A)$, and $w_3(C) > w_3(A) > w_3(B)$ (for example, $w_1(A,B,C) = (10,8,5)$, $w_2(A,B,C) = (5,10,8)$, $w_3(A,B,C) = (8,5,10)$). Winning coalitions are given by $\mathcal{W}_A = \{X \in \mathcal{C} : 3 \in X\}$, $\mathcal{W}_B = \{X \in \mathcal{C} : 1 \in X\}$, $\mathcal{W}_A = \{X \in \mathcal{C} : 2 \in X\}$ (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 are no dynamically stable states 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 every single transition in every 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 2, 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 3, 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 $w_2(C) > w_2(A)$. Therefore, $\phi(C) = B$, and by Axiom 2, $\phi(B) = B$. But we have $A \succ_B B$ and $\phi(A) = A$; this means, by Axiom 3, that $\phi(B) = B$ cannot hold. This contradiction shows that with these preferences, there is no mapping ϕ that satisfies Axioms 1–3.

Example 6 ***(Nonexistence without Assumption 2(b))*** There are 3 players, $\mathcal{I} = \{1,2,3\}$, and 4 states, $\mathcal{S} = \{A,B,C,D\}$. Players' preferences satisfy $w_1(A) > w_1(B) > w_1(C) > w_1(D)$, $w_2(B) > w_2(C) > w_2(A) > w_2(D)$, and $w_3(C) > w_3(A) > w_3(B) > w_3(D)$ (for example, $w_1(A,B,C,D) = (10,8,5,4)$, $w_2(A,B,C,D) = (5,10,8,4)$, $w_3(A,B,C,D) = (8,5,10,4)$). 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 voting rule, while in state D there is majority voting rule). We then have $A \succ_D D$, $A \succ_D D$, $A \succ_D D$ and $A \succ_D B$, $B \succ_D C$, $C \succ_D A$, 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 any other state. 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.

We 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 3, 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 Axioum 3 we cannot have $\phi(D) = A$. This contradiction proves that there does not exist mapping ϕ that satisfies Axioms 1–3.

Example 7 ***(Multiple Equilibria without Assumption 3)*** There are 2 players, $I = \{1, 2\}$, and 3 states, $S = \{A, B, C\}$. Players' preferences satisfy $w_1(A) > w_1(B) > w_1(C)$, $w_2(B) > w_2(A) > w_2(C)$ (for example, $w_1(A, B, C) = (5, 3, 1)$, $w_2(A, B, C) = (3, 5, 1)$). 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). Then 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 $B \succ_C A$).

One can easily 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 one maps state C to state A, and the second one maps state C to state A. It is straightforward to verify that ϕ_1 and ϕ_2 satisfy Axioms 1–3, and also that no other mapping satisfies these Axioms. Note that the sets of stable states under these two mappings satisfy $\mathcal{D}_{\phi_1} = \{A, B\} = \mathcal{D}_{\phi_2}$, as they should according to Theorem 1.

Proofs of Propositions From Section 6

Proof of Proposition 1 (Part 1) Take $m_{s_k} = (k+1)/2$ if k is odd and $m_s = k/2$ if k is even. Evidently, for any of the rules $\mathcal{W}_{s_k}^{maj}$, $\mathcal{W}_{s_k}^{med}$, or $\mathcal{W}_{s_k}^{l_k}$ where $k/2 < l_k \leq k$ for all k, m_{s_k} is a quasi-median voter and, moreover, the sequence $\{m_{s_k}\}_{k=1}^{N}$ is monotonically increasing.

(Part 2) In all cases $\mathcal{W}_{s_k}^{maj}$, $\mathcal{W}_{s_k}^{med}$, or $\mathcal{W}_{s_k}^{l_k}$ where $k/2 < l_k \leq k$, Assumption 1 trivially holds. From part 1 it follows that Theorem 3 (part 1) is applicable, so Assumption 2(a,b) holds. Finally, Assumption 2b* follows from 5, as Theorem 3 (part 3) is applicable in this case.

(Part 3) In an odd-sized club s_k, median voter is a single person $(k+1)/2$, and in the case of majority voting, we have $s_l \succ_{s_k} s_k$ if and only if $w_{(k+1)/2}(s_l) > w_{(k+1)/2}(s_k)$ because of the single-crossing condition. In either case, if s_l and s_j are two different clubs, player $(k+1)/2$ is not indifferent between them by Assumption 5. This implies that either $s_l \succ_{s_k} s_j$ of $s_j \succ_{s_k} s_l$ for any s_j and s_l, which completes the proof. ∎

Proof of Proposition 2. (Part 1) Assumption 1 holds in each club s_k, because the voting rule is simple majority. To show that Assumption 2(a) holds, we notice that it is impossible to have $s_l \succ_{s_k} s_k$ for $l > k$, because all members of s_k prefer s_k to s_l. Therefore, any cycle that we hypothesize to exist will break at its smallest club. To show that Assumption 2(b) holds, take any club $s = s_k$. The set of clubs $\{s_l\}$ that satisfy $s_l \succ_{s_k} s_k$ is the set of clubs that satisfy $k/2 < l \leq k$. Hence, for any clubs s_l, s_m with $l < m$ that satisfy $s_l \succ_{s_k} s_k$ and $s_m \succ_{s_k} s_k$ we have $s_l \succ_{s_k} s_m$: indeed, players $i \in \{1, \ldots, l\}$ which form a simple majority will prefer s_l to s_m, as they are included in both clubs, but prefer the smaller one. Therefore, $s_m \succ_{s_k} s_l$ is impossible for $l < m$, which proves that Assumption 2(b) holds. Likewise, $s_m \succeq_{s_k} s_l$ is impossible, so the first part of Assumption 2b* holds as well.

Let us now take $s_l \succ_{s_k} s_k$ and $s_m \not\succ_{s_k} s_k$. This means $k/2 < l \leq k$, and either $m \leq k/2$ or $m \geq k$. If $m \leq k/2$, then the set of members of club s_k who prefer s_m to s_l is $\{1, \ldots, m\}$: those who belong to s_l but not to s_m prefer s_l, while those who do not belong to either of s_m and s_l are indifferent. So, players only players in s_m may strictly prefer s_m to s_l. But they do not constitute at least half of the club in s_k, so $s_m \not\succ_{s_k} s_l$. Consider the second case, $m \geq k$. But then all players in s_l (i.e., a majority) will prefer s_l to s_m, and therefore $s_m \not\succ_{s_k} s_l$. We have proved that Assumption 2b* holds.

Finally, to show that Assumption 3 holds, take $s = s_k$, s_l and s_m such that $s_l \succ_{s_k} s_k$, $s_m \succ_{s_k} s_k$, and $s_l \nsim s_m$. Without loss of generality assume $l < m$. But then $s_l \succ_{s_k} s_m$, since all players from s_l prefer s_l, and they form a majority in s_k. This proves that Assumption 3 holds.

(Part 2) Monotonic median voter property holds, since we can take m_{s_k} to be player $k/2$ if k is even and $(k+1)/2$ is odd; clearly, $\{m_{s_k}\}_{k=1}^{N}$ is an increasing sequence of quasi-median voters. To show that the single-crossing condition holds, take $i, j \in \mathcal{I}$ such that $i < j$ and $s_k, s_l \in \mathcal{S}$ with $k < l$. Suppose $w_i(s_l) > w_i(s_k)$. This is possible if $i \in s_l$ but $i \notin s_k$ or $i \notin s_k, s_l$. In either case, $i \notin s_k$, and therefore $j \notin s_k$. But then $w_j(s_l) > w_j(s_k)$. Suppose now that $w_j(s_l) < w_j(s_k)$; this means that $j \in s_k, s_l$. But then $i \in s_k, s_l$, and therefore $w_i(s_l) < w_i(s_k)$. This establishes that the single-crossing condition holds.

(Part 3) Notice that it is never possible that $s_l \succ_{s_k} s_k$ if $k < l$. We can therefore start with smaller clubs. Club s_1 is stable and $1 = 2^0$. Suppose we proved the statement for $j < k$ and now consider club s_k. If $\log_2 k \notin \mathbb{Z}$, then club s_j for $j = 2^{\lfloor \log_2 k \rfloor}$ is stable and contains more than half members of s_k. Hence, s_k is unstable. Conversely, if $\log_2 k \in \mathbb{Z}$, then the only clubs we know to be stable do not contain more than $k/2$ members, so s_k is stable. This proves the induction step.

(Part 4) If $\log_2 k \in \mathbb{Z}$, then $2^{\lfloor \log_2 k \rfloor} = k$, and the statement follows from part 3. If $\log_2 k \notin \mathbb{Z}$, then $s_{2^{\lfloor \log_2 k \rfloor}}$ is the only club which is preferred to s_k by a majority (other stable clubs are either larger than s_k or at least twice as small as $s_{2^{\lfloor \log_2 k \rfloor}}$, i.e., more than two times smaller than s_k). The result follows. ■

Proof of Proposition 3. **(Part 1)** Assumption 1 follows from $b > N/2$. Therefore, Theorem 3 applies and Assumption 2(a,b) and 2b* are satisfied.

(Part 2) By part 1, Theorem 1 is applicable. The result immediately follows.

(Part 3) We prove this result for constitutions and voting rules simultaneously. By definition, a voting rule (constitution) (a,b) is self-stable (or self-stable, in terminology of Barbera and Jackson) if $|i \in \mathcal{I} : w_i(a') > w_i(a)| < b$ for all feasible a'. But this is equivalent to $(a',b') \not\succ_{(a,b)} (a,b)$ for all (a,b), which is the definition of myopic stability. By Corollary 1, any myopically stable state is dynamically stable, but not vice versa, which establishes the result.

(Part 4) In light of part 3 we only need to prove that any dynamically stable state is myopically stable. Take any constitution (a,b) which is not myopically stable; 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 nonempty. 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 3 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 4. **(Part 1)** Assumption 1 immediately follows from (18) and that $\alpha \geq 1/2$. To prove that Assumption 2(a) holds, it suffices to notice that $Y \succ_X X$ is impossible if $\gamma_Y > \gamma_X$, so any cycle would break at the least powerful coalition in it (which is unique because of genericity). Similarly, to prove that Assumption 2(b) holds, one can notice that if $Y \succ_X X$ and $Z \succ_X X$, then $\gamma_Y > \gamma_Z$ implies $Z \succ_X Y$, and thus $Y \not\succ_X Z$: indeed, all players in Z prefer Z to Y, and they form a winning coalition in X, for if they did not, $Z \succ_X X$ would be impossible. Again, this means that any cycle would break at the least powerful coalition in it. One can similarly show that Assumption 3 holds: is proved likewise: if $Y \succ_X X$ and $Z \succ_X X$, then, by genericity, $X \nsim Y$ implies $\gamma_Y \neq \gamma_Z$. Without loss of generality, $\gamma_Y > \gamma_Z$, and in this case $Z \succ_X Y$. Hence, Assumption 3 is satisfied.

(Part 2) Let us perturb players' payoffs so that if $i \notin X$, then $w_i(X) = \varepsilon\gamma_X$ where $\varepsilon > 0$ is small. After this perturbation, Assumptions 1, 2 and 3 still hold, as the proofs from part 1 are still valid. The first part of Assumption 2b* follows, for if a corresponding $\succeq$-cycle existed, then by genericity we would get a $\succ$-cycle which is ruled out by Assumption 2(b). To show that the second part of Assumption 2b* holds, take $Y \succ_X X$ and $Z \not\succ_X X$. This implies $\alpha\gamma_X < \gamma_Y < \gamma_X$ and either $\gamma_Z \leq \alpha\gamma_X$ or $\gamma_Z > \gamma_X$. If $\gamma_Z \leq \alpha\gamma_X$, all players who are not in Z prefer Y to Z: this is obviously true for the part that belongs to Y, while if a player is neither in Y nor in Z, this is true because of the perturbation we made, for in this case $\gamma_Y > \alpha\gamma_X \geq \gamma_Z$. Since players in Z do not form a winning coalition in this case, we have $Z \not\succ_X Y$. Consider the second case where $\gamma_Z > \gamma_X$; then all players in Y prefer Y to Z, since $\gamma_Y < \gamma_Z$. This means that $Y \succ_X Z$ and thus $Z \not\succ_X Y$. This proves that Assumption 2b* holds, which completes the proof. ■

The Relationship Between $\mathcal{D}$, von Neumann-Morgenstern Stable Set, and Chwe's Largest Consistent Set

The following definitions are from Chwe (1994) and von Neumann and Morgenstern (1944).

Definition 11 ***(Consistent Sets)*** *For any $x, y \in \mathcal{S}$ and any $X \in \mathcal{C}$, define relation $\rightarrow_X$ by $x \rightarrow_X y$ if and only if either $x = y$ or $x \neq y$ and $X \in \mathcal{W}_x$.*

1. *We say that state x is* ***directly dominated*** *by y (and write $x < y$) if there exists $X \in \mathcal{C}$ such that $x \rightarrow_X y$ and $x \prec_X y$, where we write $x \prec_X y$ as a shorthand for $w_i(x) < w_i(y)$ for all $i \in X$.*

2. *We say that state x is* ***indirectly dominated*** *by y (and write $x \ll y$) if there exist $x_0, x_1, \ldots, x_m \in \mathcal{S}$ such that $x_0 = x$ and $x_m = y$ and $X_0, X_1, \ldots, X_{m-1} \in \mathcal{C}$ such that $x_j \rightarrow_{S_j} x_{j+1}$ and $x_j \prec_{S_j} y$ for $j = 0, 1, \ldots, m-1$.*

3. *A set $S \subset \mathcal{S}$ is called* ***consistent*** *if $x \in S$ if and only if $\forall y \in \mathcal{S}, \forall X \in \mathcal{C}$ such that $x \rightarrow_X y$ there exists $z \in S$, where $y = z$ or $y \ll z$, such that $x \not\prec_X z$.*

Definition 12 ***(von Neumann-Morgenstern's Stable Set)*** *A set of states $X \subset \mathcal{S}$ is* ***von Neumann-Morgenstern stable*** *if it satisfies the following properties:*

1. ***(Internal stability)*** *For any $x, y \in X$ we have $y \not\succ_x x$;*

2. ***(External stability)*** *For any $x \in \mathcal{S} \setminus X$ there exists $y \in X$ such that $y \succ_x x$.*

Proposition 6 *Suppose Assumptions 1 and 2 hold. Then:*

1. *The set of stable states $\mathcal{D}$ is the unique von Neumann-Morgenstern stable set;*

2. *$\mathcal{D}$ is the largest consistent set;*

3. *Any consistent set is either $\mathcal{D}$ or any subset of the set of exogenously stable states (and vice versa, all such sets are consistent).*

Proof. (Part 1) We take the sequence of states $\left\{\mu_1, ..., \mu_{|\mathcal{S}|}\right\}$ satisfying (7). Suppose that set of states $\mathcal{X}$ is von Neumann-Morgenstern stable; let us prove that $\mathcal{X} = \mathcal{D}$. Clearly, $\mu_1 \in \mathcal{X}$, since $\mu_k \not\succ_{\mu_1} \mu_1$ for any state μ_k. Now suppose that we have proved that $\mathcal{X} \cap \left\{\mu_1, \ldots, \mu_{k-1}\right\} = \mathcal{D} \cap \left\{\mu_1, \ldots, \mu_{k-1}\right\}$ for some $k \geq 2$; let us prove that $\mu_k \in \mathcal{X}$ if and only if $\mu_k \in \mathcal{D}$. From Theorem 1 it follows that it suffices to prove that $\mu_k \in \mathcal{X}$ if and only if $\mathcal{M}_k = \varnothing$. Suppose first that $\mathcal{M}_k \neq \varnothing$; then, since $\mathcal{M}_k = \mathcal{X} \cap \left\{\mu_1, \ldots, \mu_{k-1}\right\}$ by construction, we have that $\mu_l \succ_{\mu_k} \mu_k$ for some $l < k$ such that $\mu_l \in \mathcal{X}$. Hence, if $\mu_k \in \mathcal{X}$, then internal stability property would be violated, and therefore $\mu_k \notin \mathcal{X}$. Now consider the case where $\mathcal{M}_k = \varnothing$. This means that $\mathcal{X} \cap \left\{\mu_1, \ldots, \mu_{k-1}\right\} = \varnothing$, and therefore there does not exist $\mu_l \in \mathcal{X}$ such that $l < k$ and $\mu_l \succ_{\mu_k} \mu_k$. But by (7), $\mu_l \not\succ_{\mu_k} \mu_k$ whenever $l > k$. Hence, for any $\mu_l \in \mathcal{X}$ such that $l \neq k$ we have $\mu_l \not\succ_{\mu_k} \mu_k$, and therefore $\mu_k \in \mathcal{X}$, for otherwise external stability condition would be violated. This proves the induction step, and therefore completes the proof that $\mathcal{X} = \mathcal{D}$.

(Part 2) It is obvious that for any $x, y \in \mathcal{S}$, $x < y$ implies $x \ll y$. In our setup, however, the opposite is also true, so $x < y$ if and only if $x \ll y$. To see this, suppose that $x \ll y$; take a sequence of states and a sequence of coalitions as in Definition 11. Let $k \geq 0$ be lowest number such that $x_{k+1} \neq x$. This means that $x \to_{X_k} x_{k+1}$ (because $x_k = x$) and $\forall i \in X_k : w_x(i) < w_y(i)$. By definition, $x < y$; note also that $X_k \in \mathcal{W}_x$, since $x \neq x_{k+1}$.

To show that set $\mathcal{D}$ is consistent, consider some mapping ϕ that satisfies Axioms 1–3. Take any $x \in \mathcal{D}$, and then take any $y \in \mathcal{S}$ and any $X \in \mathcal{C}$ such that $x \to_X y$. Let $z = \phi(y)$; then, as follows from Axiom 1, either $z = y$ or $y \ll z$. Now consider two possibilities: $x = y$ and $x \neq y$. In the first case, $x = y \in \mathcal{D}$, so $z = y = x$. Since X is nonempty, property $\exists i \in X : w_x(i) \geq w_z(i)$ is satisfied. Now suppose that $x \neq y$; then $X \in \mathcal{W}_x$. On the other hand, $z \in \mathcal{D}$. But it is impossible that $z \succ_x x$, since both x and z are stable (otherwise, Axiom 1 would be violated for mapping ϕ), hence, in this case, $\exists i \in X : w_i(x) \geq w_i(z)$, too.

Now take some $x \notin \mathcal{D}$. We need to show that there exist $y \in \mathcal{S}$ and $X \in \mathcal{C}$ such that $x \to_X y$ and for any $z \in \mathcal{D}$ which satisfies that either $z = y$ or $y \ll z$, we necessarily have $\forall i \in X : w_i(x) < w_i(z)$. Take $y = \phi(x)$ and $X = \{i \in \mathcal{I} : w_i(x) < w_i(y)\} \in \mathcal{W}_x$; then $x \to_X y$. Note that it is impossible that for some $z \in \mathcal{D}$ we have $y \ll z$, for then $y < z$, and therefore $z \succ_y y$, which would violate Axiom 1. Therefore, any $z \in \mathcal{D}$ such that either $z = y$ or $y \ll z$

must satisfy $z = y$. But then, by our choice of X, we have $\forall i \in X : w_i(x) < w_i(z)$. This proves that $\mathcal{D}$ is indeed a consistent set.

To show that $\mathcal{D}$ is the largest consistent set, suppose, to obtain a contradiction, that the largest consistent set is $S \neq \mathcal{D}$. Since $\mathcal{D}$ is consistent, we must have $\mathcal{D} \subset S$. Consider sequence $\left\{\mu_1, ..., \mu_{|\mathcal{S}|}\right\}$ satisfying (7), and among all states in $S \setminus \mathcal{D} \neq \varnothing$ pick state $x = \mu_k \in S \setminus \mathcal{D}$ with the smallest number, i.e., such that if $\mu_l \in S \setminus \mathcal{D}$, then $l \geq k$. We now show that, according to the definition of a consistent set, $x \notin S$, which would contradict the assertion that state S is consistent. Take some mapping ϕ that satisfies Axioms 1–3. Now let $y = \phi(x) \in \mathcal{D}$ and $X = \{i \in \mathcal{I} : w_i(x) < w_i(y)\} \in \mathcal{W}_x$; then $x \rightarrow_X y$ and, since $x \notin \mathcal{D}$, $y \neq x$, which by (7) implies that $y = \mu_l$ for $l < k$. Now if for some $z \in S$ we have $y \ll z$, then $y < z$, and hence $z \succ_y y$, which implies $z = \mu_j$ for some $j < l < k$. But then $z \notin S \setminus \mathcal{D}$, and therefore $z \in \mathcal{D}$. However, it is impossible that $y, z \in \mathcal{D}$ and $z \succ_y y$, as this would violate Axiom 1. Therefore, if for some $z \in S$ either $z = y$ or $y \ll z$, then in fact $z = y$. But for such z, we do have $\forall i \in X : w_i(x) < w_i(z)$, by construction of X. We get a contradiction, since by definition of a consistent set $x \notin S$, while we picked $x \in S \setminus \mathcal{D}$. This proves that $\mathcal{D}$ is the largest consistent set.

(Part 3) By part 2, if S is a consistent set, then $S \subset \mathcal{D}$. Suppose that $S \neq \mathcal{D}$, but S includes a state which is not exogenously stable. Suppose $x \in S$ is not exogenously stable and $y \in \mathcal{D} \setminus S$; then $x \rightarrow_X y$ for some $X \in \mathcal{W}_x$. Since $x \in S$, there exists $z \in S$ where either $z = y$ or $y \ll z$, such that $\exists i \in X : w_i(x) \geq w_i(z)$. But $y \in \mathcal{D} \setminus S$, and hence $y \ll z$, which implies, as before, $y < z$ and $z \succ_y y$. However, this is impossible, since $y, z \in \mathcal{D}$. This contradiction proves that if $S \neq \mathcal{D}$, S may not include any state which is not exogenously stable.

Consider, however, any S which consists of exogenously stable states only. Take any $x \in S$. If $y \in \mathcal{S}$ and $X \in \mathcal{C}$ are such that $x \rightarrow_X y$, then $x = y$. In that case, we can take $z = y \in S$ and find that condition $\exists i \in X : w_i(x) \geq w_i(z)$ trivially holds. Now take any $x \notin S$. Consider two possibilities. If state x is exogenously stable, then take $X = \mathcal{I}$ and $y = x$; then $x \rightarrow_X y$. If for some $z \in S$ we had $y \ll z$, then, in particular, $y \rightarrow_Y z$ for some $Y \in \mathcal{C}$, which is incompatible with $z \neq y$; at the same time, $z = y$ is impossible, as $z \in S$ and $y = x \notin S$. This means that for this y there does not exist $z \in S$ such that either $z = y$ or $y \ll z$, and therefore $x = y$ should not be in S. Finally, suppose that x is not exogenously stable. Again, consider mapping ϕ satisfying Axioms 1–3 and take $y = \phi(x)$ and $X = \{i \in \mathcal{I} : w_i(x) < w_i(y)\} \in \mathcal{W}_x$; then $x \rightarrow_X y$. By the same reasoning as before, if for some $z \in S$ either $z = y$ or $y \ll z$, then $z = y$, because $y \ll z$ would imply $z \succ_y y$ for $y, z \in \mathcal{D}$. But for such z, we have $\forall i \in X : w_i(x) < w_i(z)$ by construction of X. This proves that S is indeed a consistent set, which completes the proof. ■