Dynamics and Stability of Constitutions, Coalitions, and Clubs

Daron Acemoglu, Georgy Egorov, and Konstantin Sonin

Web Appendix

Examples, Applications and Additional Results

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 6 ***(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 7 ***(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*

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

Examples

Example 3 ***(Nonexistence if*** β ***is not close to*** 1***)*** There are 4 players, $\mathcal{I} = \{1,2\}$, and 4 states, $S = \{A, B, C, D\}$. Players' preferences are given by: $w_1(A,B,C,D) = (90,70,60,5)$, $w_2(A,B,C,D) = (5,50,40,30)$, $w_3(A,B,C,D) = (25,50,40,30)$, $w_4(A,B,C,D) = (25,25,40,30)$. Winning coalitions are defined as follows: in states A, B, C, player 1 is the dictator, while in state D, players 2, 3, 4 make decisions by majority voting. It is straightforward to show that Assumptions 1, 2. (The only condition to be checked is that Assumption 2(b) holds for state $s = D$, and this follows from the fact that $B \succ_D D$, $C \succ_D D$, but $A \not\succ_D B$ and $A \succ_D C$.) Suppose, however, that the discount factor β is not close to 1, say, $\beta = 1/2$; there are either no transaction costs or small transaction costs. The protocol at any state is $\pi = (A, B, C, D)$ (with the current state skipped).

Suppose that there exists an equilibrium in pure strategies. Given that player 1 is the dictator in states A, B, C, we immediately get that if the game is at state A, no transition will happen, and if the game is at either B or C, then there will be an immediate transition to A. Consider now what will happen if the state is D. Consider all four possibilities: no transition, transition to A, transition to B, and transition to C.

If there is no transition in equilibrium and alternative C is voted, it will be accepted as players 3 and 4 will support it (even though they prefer C, but not A where C ultimately leads, to D, they still prefer the path from C to A to staying in D. This also means that it will be proposed along the equilibrium path. Hence, D cannot be stable.

If there is transition to A in equilibrium, then consider the last voting where, if reached on or off the equilibrium path, A will be proposed. All of the players 2, 3, 4 will prefer to vote against this alternative, as any other transition, as well as staying in D, will lead to a higher payoff for each of them. Hence, transition to A cannot happen in equilibrium.

Suppose that there is transition to state B. Then again, players 2 and 4 would prefer to stay in D, even though this may mean transiting to B in the next period. Voting against B will, however, lead to voting on C, so we need to verify that C will be rejected at this voting. Accepting C will lead to C and then to A, while rejecting will lead to D, and then (as transition to B happens in equilibrium) to B and then to A. The latter is preferred by players 2 and 3, which means that C will be rejected if B is rejected. Consequently, players 2 and 4 are better off voting against B, which means that transition to B may not happen in equilibrium.

Finally, suppose that transition to state C happens in equilibrium. If so, when alternative B is $v$oted, players 2 and 3 will support B, as they prefer to transit to A through B rather than through C. This implies that transition to C cannot happen in equilibrium either. In all cases, we have reached a contradiction, which means that there is no pure-strategy MPE in this case.

Example 4 ***(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 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 5 ***(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 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 stage 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 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 6 ***(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 transaction 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 7 ***(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 thus 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 8 ***(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.

Proof of Lemma 1

(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 $\{i \in \mathcal{I} : w_i(x) > w_i(y)\} \in \mathcal{W}_z$ (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

$$m_s = \max_{x \in \mathcal{S} : x \leq s} \min_{m \in M_x} m. \tag{B1}$$

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 (B1). 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 (B1) is increasing, part 3 follows.

Transaction Cost and Discount Factor

In the proof of Theorem 2 in Appendix A, the two conditions that the discount factor β has to satisfy are given by (A3) and (A4). Recall that in footnote 17, we defined $\bar{\varepsilon} = \max_{i \in \mathcal{I}, x \in \mathcal{S}} |w_i(x) - \tilde{w}_i|$. Suppose that $\bar{\varepsilon}$ increases, which means that at least for one individual i, payoff during transition, $\tilde{w}_i$, decreases. This makes both (A3) and (A4) harder to satisfy for a given β, but both conditions hold for some higher β. Consequently, for any $\bar{\varepsilon}$ there exists $\beta_0 < 1$ such that for $\beta > \beta_0$, Theorem 2 holds. This also implies that for any $\bar{\varepsilon} > 0$, as $\beta \to 1$, discounted payoffs are independent of transaction costs (i.e., do not depend on $\bar{\varepsilon}$).

Additional Applications

We now illustrate how the characterization results provided in Theorems 1 and 2 can be applied in a number of political economy environments considered in the literature. We show that in some of these environments we can simply appeal to Theorem 4. Nevertheless, we will also see that the conditions in Theorem 4 are more restrictive than those stipulated in Theorems 1 and 2. Thus, when Theorem 4 does not apply, Theorems 1 and 2 may still be applied directly.

Voting in Clubs

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:

$$\text{for all } l > k \text{ and } j > i, \;\; w_j(s_l) - w_j(s_k) > w_i(s_l) - w_i(s_k), \tag{B2}$$

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 the environment introduced in Roberts (1999) is a special case of our environment, 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

$$\text{for all } l > k \text{ and } j > i, \qquad \begin{aligned} w_i(s_l) > w_i(s_k) &\implies w_j(s_l) > w_j(s_k), \text{ and} \\ w_j(s_k) > w_j(s_l) &\implies w_i(s_k) > w_i(s_l). \end{aligned} \tag{B3}$$

Clearly, (B2) implies (B3) (but not vice versa). In addition, Roberts's two voting rules can be represented by the following sets of winning coalitions:

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

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

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 5. Let us also assume that Assumption 6 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 4 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 is important, however, to emphasize that while our model nests Roberts' environment as a special case, the characterization of MPE is obtained here, unlike Roberts's paper, only under the assumption of transaction costs and a sufficiently large discount factor.

It can also be verified that Theorem 4 applies with considerably more general voting rules (e.g., with different degrees 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 4 establishes the following proposition.

Proposition 4 *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 5 is satisfied.

(ii) Suppose that preferences satisfy (B3) and Assumption 6. Then Assumptions 2(a,b) 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 (up to payoff-equivalence) as a function of the initial state (club).

Proof of Proposition 4. **(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 4 (part 1) is applicable, so Assumption 2(a,b) holds.

(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 6. 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.

This proposition shows that a sharp characterization of dynamics of clubs and the set of stable clubs can be obtained easily by applying Theorem 4 to Roberts's original model or to various generalizations. Another generalization, not stated in Proposition 4, 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.

Inefficient Inertia and Lack of Reform

We now provide a more detailed example capturing the main trade-offs discussed as motivation 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 3. 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.

We can take this line of argument even further. 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 $\mathcal{W}_{d1} = \mathcal{W}_{d2} = \{X \in \mathcal{C} : P \in X\}$ and

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

so that P prefers "extensive" redistribution. Given the fact that $\mathcal{W}_{d1} = \mathcal{W}_{d2} = \{\{P\}, \{E, P\}\}$, 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, 2006, for examples in which the middle class may have played such a role in the process of democratization).

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).

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).[26]

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 $\gamma_X \equiv \sum_{j \in X} \gamma_j$. Suppose also that payoffs are given by

$$w_i(X) = \begin{cases} \gamma_i / \gamma_X & \text{if } i \in X \\ 0 & \text{if } i \notin X \end{cases} \tag{B4}$$

for any $i \in \mathcal{I}$ and any $X \in \mathcal{C} \equiv \mathcal{S}$.[27] The restriction to (B4) 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

$$\mathcal{W}_X = \left\{ Y \in \mathcal{C} : \sum\nolimits_{j \in Y \cap X} \gamma_j > \alpha \sum\nolimits_{j \in X} \gamma_j \right\}. \tag{B5}$$

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:

$$\gamma_X = \gamma_Y \text{ only if } X = Y. \tag{B6}$$

[26] In Acemoglu, Egorov and Sonin (2008), not all transitions are allowed. In particular, the focus is on a game of "eliminations" from ruling coalitions in nondemocracies, so that once a particular individual is eliminated, he can no longer be part of future ruling coalitions (either because he is "killed," permanently exiled, or is permanently excluded from politics by other means). In Web Appendix, we allow for restrictions on feasible transitions and show how Proposition 5 can be generalized to cover the case of political eliminations considered in Acemoglu, Egorov, and Sonin (2008).

[27] 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 (B4) is adopted to simplify the discussion here.

The following proposition can now be established.

Proposition 5 *Consider the environment in Acemoglu, Egorov, and Sonin (2008). Then there exists an arbitrarily small perturbation of payoffs such that Assumptions 1, 2(a,b), and 3 are satisfied. Then Theorem 1 and Theorem 2 apply and characterize the stable states.*

Proof of Proposition 5. Let us perturb players' payoffs so that if $i \notin X$, then $w_i(X) = \varepsilon\gamma_X$ where $\varepsilon > 0$ is small. Assumption 1 immediately follows from (B5) 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, notice that if a $\succeq$-cycle exists, it is by genericity a $\succ$-cycle. But 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. Now, 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 $Z \not\succ_X Y$. One can similarly show that Assumption 3 holds: 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 $Z \succ_X Y$. This completes the proof.

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 $n_0 = n_1 < j \leq n_2 < n_3$,

$$w_k(s_{n_0}) = w_k(s_{n_1}) < w_k(s_{n_3}) < w_k(s_{n_2}). \tag{B7}$$

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

$$\mathcal{W}_{s_k} = \{X \in \mathcal{C} : |X \cap s_k| > k/2\}. \tag{B8}$$

It is straightforward to verify that this environment satisfies Assumptions 1, 2(a,b), and 3.[28] 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

$$\phi(s_k) = s_{2^{\lfloor \log_2 k \rfloor}}, \tag{B9}$$

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 6 *In the elite club example considered above with preferences given by (B7) and set of winning coalitions given by (B8), the following results hold.*

1. *Assumptions 1, 2(a,b), and 3 hold.*
2. *If, instead of (B7), 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 $n \in \mathbb{Z}^+$.*
4. *The unique mapping ϕ that satisfies Axioms 1–3 is given by (B9).*

[28] 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 4 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).

Proof of Proposition 6. **(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 $k/2 < l < k$. Hence, for any clubs s_l, s_m with $l < m$ that satisfy $s_l \succeq_{s_k} s_k$ and $s_m \succeq_{s_k} s_k$ 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, $s_m \succeq_{s_k} s_l$ is impossible for $l < m$. 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 2(b) 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.

Stable Voting Rules and Constitutions

Another interesting model that can be analyzed using Theorem 4 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

$$w_i\left[(a, b)\right] = w_i\left(a\right). \tag{B10}$$

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

$$\mathcal{W}_{(a.b)} = \{X \in \mathcal{C} : |X| \geq b\} \tag{B11}$$

(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 (B10) and (B11). 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 $w_i\left(a\right)$, and thus $w_i\left[(a, b)\right]$, satisfies the single-crossing condition in Definition 3. This enables us to apply Theorem 4 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. The following proposition states the relation between self-stable constitutions and dynamically stable sets.

Proposition 7 *Consider the above-described environment and assume that preferences satisfy single-crossing condition and Assumption 6 holds. Then:*

1. *Assumptions 1, 2(a,b) 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. *The set of self-stable constitutions coincides with the set of dynamically stable states.*

Proof of Proposition 7. (Part 1) Assumption 1 follows from $b > N/2$. Therefore, Theorem 4 applies and Assumption 2(a,b) are satisfied.

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

(Part 3) By definition, a constitution (a, b) is self-stable 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). By (5) we obtain that $\phi_c[(a, b)] = (a, b)$, i.e., (a, b) is ϕ_c-stable. Hence, a self-stable constitution is a dynamically stable state.

Vice versa, take any dynamically stable state (a, b). Suppose, to obtain a contradiction, that (a, b) is not a self-stable constitution; let us prove that then $\phi_c[(a, b)] \neq (a, b)$. Consider the set of constitutions $\mathcal{Q} = \{(a', b')\}$ such that $(a', b') \succ_{(a,b)} (a, b)$; since (a, b) is not self-stable, 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.

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 legislative bargaining.[29]

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, 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 $\{1,2\}$ forms, party 1, by virtue of its greater number of seats, can sideline party 2 and rule by itself. Let us introduced the

[29]See, for example, David Baron and John Ferejohn (1986), Austen-Smith and Banks (1988), Baron (1991), Jackson and Boaz Moselle (2002), and Peter Norman (2002) for models of legislative bargaining. The recent paper by Daniel Diermeier and Pohan Fong (2011) that studies legislative bargaining as a dynamic game without commitment also raises a range of issues related to our general framework here.

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 that the instability of states that can be reached from a state s contributes to the stability of state s.

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., James Fearon, 1996, 2004 and Robert 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, 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.

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

$$w_i(s_j) = \mathbb{E}\left[\left(1 - \tau_{s_j}\right) A_i + G_{s_j}\right], \tag{B12}$$

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 (B12) 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,

$$\mathbb{E}\tau_{s_{l+1}} > \mathbb{E}\tau_{s_l} \tag{B13}$$

is sufficient for single-crossing. Note that individual i, when determining the tax rate in s_l, would maximize $(1 - \tau) A_i + h\left(\tau \sum_{m=1}^{l} A_m\right)$. This implies that individual i would choose τ_i

such that

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

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, (B13) holds and we can apply Theorem 4 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).

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

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

Definition 8 ***(Consistent Sets)*** *For any $x, y \in \mathcal{S}$ and any $X \in \mathcal{C}$, define relation $\to_X$ by $x \to_X y$ if and only if either $x = y$ or $x \neq y$ and $X \in \mathcal{W}_x$.*

Definition 9

1. *We say that state x is* ***directly dominated*** *by y (and write $x < y$) if there exists $X \in \mathcal{C}$ such that $x \to_X y$ 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 $x_j \to_{S_j} x_{j+1}$ 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 $x \to_X y$ there exists $z \in S$, where $y = z$ or $y \ll z$, such that $x \not\prec_X z$.*

Definition 10 ***(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 8 *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 of Proposition 8. (Part 1) We take the sequence of states $\left\{\mu_1, ..., \mu_{|\mathcal{S}|}\right\}$ satisfying (A1). 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 $\mathcal{X} \cap \{\mu_1, \ldots, \mu_{k-1}\} = \mathcal{D} \cap \{\mu_1, \ldots, \mu_{k-1}\}$ 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 $\mathcal{M}_k = \mathcal{X} \cap \{\mu_1, \ldots, \mu_{k-1}\}$ 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 $\mathcal{X} \cap \{\mu_1, \ldots, \mu_{k-1}\} = \varnothing$, 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 (A1), $\mu_l \not\succ_{\mu_k} \mu_k$ 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 8. 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 $x \rightarrow_X y$ 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 $x \rightarrow_X y$. 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 (A1), 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 (A1) 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.

Additional References

Acemoglu, Daron. 2003. "Why Not a Political Coase Theorem? Social Conflict, Commitment, and Politics." *Journal of Comparative Economics*, 31: 620-652.

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

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.

Diermeier, Daniel and Pohan Fong. 2011. "Legislative Bargaining with Reconsideration." *Quarterly Journal of Economics*, forthcoming.

Fearon, James. 1998. "Commitment Problems and the Spread of Ethnic Conflict." In *The International Spread of Ethnic Conflict: Fear, Diffusion, and Escalation*, ed. David Lake and Donald Rothchild. Princeton, NJ: Princeton University Press.

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

Jackson, Matthew O. and Boaz Moselle . 2002. "Coalition and Party Formation in a Legislative Voting Game." *Journal of Economic Theory*, 103(1): 49-87.

Norman, Peter. 2001. "Legislative Bargaining and Coalition Formation." *Journal of Economic Theory*, 102(2): 322-353.

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

Schwarz, Michael, and Konstantin Sonin. 2008 "A Theory of Brinkmanship, Conflicts, and Commitment." *Journal of Law, Economics, and Organization*, 24 (1): 161-183.

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