A Spatial Theory of Party Formation.*

Jon X Eguia†

New York University

January 18, 2011

Abstract

Members of an assembly that chooses policies on a series of multidimensional ideological issues have incentives to coalesce and coordinate their votes, forming political parties. If an agent has an advantage to organize a party at a lower cost, a unique party forms and the policy outcome moves away from the Condorcet winning policy, to the benefit of party members. If all agents have the same opportunities to coalesce into parties, at least two parties form. The results are robust to the consideration of an endogenous agenda, and to generalizations of the distribution of preferences.

*I thank Mik Laver, Luis Corchón and participants in the 2008 Workshop on the Political Economy of Democracy in Barcelona, the 2008 Social Choice and Welfare conference in Montreal, and seminars at U. Autònoma Barcelona, U. Essex and the London School of Economics for insightful comments and suggestions. I completed this paper while visiting the Kellogg School of Management (Northwestern University), and I am grateful for financial support from its Ford Center for Global Citizenship and its Center for Mathematical Studies.

†Email: eguia@nyu.edu. Mail: 19 West 4th Street, 2nd Floor. Dept. of Politics, NYU. New York, NY 10012.

Keywords: Party formation, coalition formation.

JEL classification: D71, D72.

Members of an assembly that makes choices by majority voting have strategic incentives to coalesce and coordinate their votes.

I present a theory of coalition formation in an assembly in which any set of members of the assembly can coalesce, by unanimous agreement, into a voting bloc. Agents who coalesce casts their votes together. A prominent application is the formation of political parties that exercise party discipline.

I consider a two dimensional choice set, and a finite set of agents, each of whom has spatial preferences characterized by an ideal point in the choice set, and a ranking of alternatives that decreases in the Euclidean distance to this point. The assembly makes a sequence of binary choices, each of them pitting an invariant status quo against an alternative. I find that if there exists an agent who enjoys an advantage and can coordinate and organize a coalition at a lower cost, then a single coalition forms around this agent. If all agents face the same cost of coalescing, at least two coalitions form in every equilibrium, and an equilibrium with exactly two coalitions exists. In this equilibrium, two coalitions form with agents with ideal points on opposite sides of the choice set.

A Condorcet winner, if it exists, is an alternative that defeats any other in pairwise comparisons. A status quo is the choice made by the assembly each time an alternative does not gather a majority of votes. If the status quo is a Condorcet winner, in traditional theories of legislative choice, the policy outcome coincides with the status quo. I assume that a Condorcet winner exists and the status quo is the Condorcet winner. I find that a subset of agents whose preferences are similar in one dimension, but diverge in a second dimension, have an incentive to coalesce, committing to vote together in the assembly, according to the preference of the majority of the bloc. If these agents form a voting bloc, they defeat the Condorcet winner and the assembly's choice moves away from the status quo, to the advantage of the members of the bloc.

This paper is complementary to other theories of party formation. Snyder and Ting (2002) describe parties as informative labels that help voters to decide how to vote; Caillaud and Tirole (1999) and (2002) focus on the role of parties as information intermediaries that select high quality candidates; Ashworth and Bueno de Mesquita (2008) relate the value of the label to the quality of the screening for appropriate candidates; Osborne and Tourky (2008) argue that parties provide economies of scale; Levy (2004) stresses that parties act as commitment devices to offer a policy platform that no individual candidate could credibly stand for and Ansolabehere, Leblanc and Snyder (in this issue) describe how each party chooses this policy platform; and Morelli (2004) notes that parties serve as coordination devices for like-minded voters to avoid splitting their votes among several candidates of a similar inclination. All these theories explain party formation as a result of the interaction between candidates and voters in elections.

In contrast, my theory explains the formation of parties within the legislature. Baron (1989) and Jackson and Moselle (2002) also study the formation of parties within an assembly. However, Baron (1989) does not consider ideological preferences, presenting instead an assembly that bargains only over a purely distributive dimension. Jackson and Moselle (2002) introduce ideological preferences, but their analysis of party formation is limited to examples in an assembly with three agents, where competition between two parties is not feasible. Baron (1989), Jackson and Moselle (2002) and Baron and Hirsch (in this special issue) seek to explain the formation of a ruling coalition that distributes pork or office-holding benefits. I seek instead to explain the incentives to form parties to affect the policy outcome in a purely ideological space of preferences. In Eguia (2010) I introduce a model of party formation with binary preferences over a binary policy space, and in Eguia (2011) I extend this model to a repeated game. In the current paper I present an alternative model with two main innovations: First, I consider a richer policy space, letting the set of feasible policies be a subset of a two dimensional vector space, making it possible to consider a continuum of ideological preferences over each of two issues. Second, I endogenize the agenda, letting an agent choose the policy proposals strategically. Diermeier and Vlaicu (2008) show that a majority of legislators who ex-ante share the same type form a party to gain control of the agenda. My theory is robust to the consideration of both an exogenous and an endogenous agenda, while allowing for agents who all have ex-ante different preferences.

A political party that exercises party discipline and functions as a voting bloc aggregates the preferences of its members so that the internal minority within the party always reverses its vote, to vote along with the majority of the party. This coordination may affect the policy outcome in a given issue, benefiting the majority, and hurting the minority. Repetition of such behavior over a sequence of decisions can benefit every member if the identity of the minority and majority within the party changes across decisions.

The gains made by a set of agents that trades votes are well known to literature on the log rolling and vote trading. Fox (2006) finds sufficient conditions for a majority of legislators to benefit if they cooperate. Carruba and Volden (2000), who after laying out their theory of log rolling, they speculate about the role of parties. They suggest that parties are perhaps coordination devices: "groups of legislators who agree to support one another's legislation and exclude others." The current manuscript pursues this idea. See as well Koford (1982), for a model in which legislators purchase and sell votes at a price from a party leader that acts as auctioneer of legislative majorities; and Stratmann (1992) for empirical evidences of logrolling in the US Congress.

The gains from the coordination of votes explain not only the formation of parties by individual legislators; in assemblies with multiple factions or parties, such as the European Parliament or the Israeli Knesset, individual parties can coalesce with others to form larger voting blocs, explaining the formation of alliances, mergers of parties, or transnational parties in the case of the European Parliament. Diermeier and Merlo (2000), Schofield and Sened (2006) and Baron, Diermeier and Fong (in this special issue) discuss the formation of a single coalition to control the government. My theory explains the formation of one or more alliances that form to influence the outcome on certain votes, even if they do not control the government.

After an illustrating example, I present the theory, and then I first show my results with an exogenous agenda, followed by similar results with an endogenous agenda, and an extension of the theory to allow for more general preferences. After a discussion of the findings, an appendix contains the proofs of all the results.

An Example

Consider an assembly with nine legislators who vote to make two choices, each of which is two dimensional. Each of the two choices consists of choosing an outcome in a policy space with two dimensions. For instance, the first choice could be a budget proposal, with government spending in the first dimension and taxation in the second dimension, and the second issue could be a new immigration law, with policy toward legal immigrants in the first dimension and policy toward illegal immigrants in the second dimension.

Assume that on both collective decision problems, the status quo is $(0,0)$, the ideal policy of agent 0 is at the status quo and, clockwise, agents 1 through 8 respectively have ideal policies at $(0,1),(1,1),(1,0),(1,-1),(0,-1),(-1,-1),(-1,0)$ and $(-1,1)$. That is, the ideal policies are distributed on a 3 by 3 grid. The utility that an agent derives from a policy outcome on any choice is linearly decreasing in the Euclidean distance from the policy outcome to the ideal policy of the agent. Utility is additive across issues. The assembly votes sequentially, considering each decision separately. The agenda on each decision puts to a vote the status quo versus a policy proposal randomly drawn from the Pareto set of policies $[-1,1]^2$. In each of the two collective decisions problems, the group choice is the policy proposal if five or more agents vote for it, and the status quo otherwise.

Note that the status quo is a Condorcet winner and defeats any proposal if agents vote their true preference. Suppose instead that agents $2,3,4$ form a voting bloc and commit to vote together, according to the their internal majority, so that if two agents agree, the third votes with them regardless of her own preference. These three agents all have an ideal policy to the right of the status quo, but they disagree on the second dimension. If the random proposal is up and slightly to the right of the status quo, agents with ideal policies at $(-1,1),(0,1),(1,1)$ and $(1,0)$ favor the proposal. Agent 4 with ideal policy $(1,-1)$ is against it, but because she belongs to the voting bloc with agents 2 and 3 who favor the proposal, she votes in favor as well. The proposal passes, agents 2 and 3 are better off and

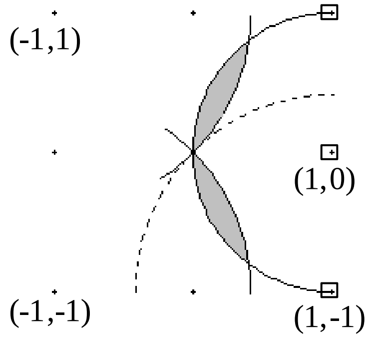

Figure 1: A voting bloc forms and changes the policy outcome.

agent 4 is worse off. Ex ante, it is equally likely that the proposal lies to the right and down from the status quo and ex post agents 3 and 4 benefit while agent 2 is worse off after $2, 3, 4$ make the proposal pass. Ex ante, both 2 and 4 benefit from the formation of a bloc.

Figure 1 illustrates the formation of a voting bloc by $2, 3$ and 4. The status quo policy is at the center of the figure, at $(0, 0)$. The agents who belong to the bloc have their ideal policies marked by a square. I depict the indifference curves of four agents through the status quo policy. The ideal policies of these agents are labelled. Three of these curves determine the two shaded areas such that if the policy proposal lies inside them, the coordination of votes inside the bloc makes the policy proposal pass. The fourth indifference curve, dashed, helps to illustrate that although the agent 4 with ideal policy at $(1, -1)$ is hurt when the proposal passes in the upper shaded area, she benefits more when it passes in the lower shaded area, so ex ante she attains a net benefit.

Note that agents with an ideal policy to the left of the status quo are ex ante worse off. They have incentives to form a second voting bloc that cancels out the first one, so that the policy outcome remains at the status quo policy $(0, 0)$. In the next section I introduce a formal game of coalition formation and sequential voting, and I show that in equilibrium at least two voting blocs form if every subset of agents has the same opportunity to coordinate and form voting blocs.

A Theory of Voting Blocs

Let there be an assembly $\mathcal{N}$ with N (odd) agents who face T decisions. Each decision consists of choosing a policy in a two dimensional policy space. On each decision t, the status quo is $(0,0)$ and a proposal $\rho^t \in \mathbb{R}^2$ is put to a vote against the status quo. The choice of the assembly on decision t is the proposal ρ^t if it gathers a simple majority of votes in favor, and the status quo $(0,0)$ otherwise. The assembly then moves on to the next decision. Legislators have circular preferences around their ideal policy, and their utility is additive across decisions, without discount. For each legislator $i \in \mathcal{N}$, let $x^i \in \mathbb{R}^2$ be the ideal policy of the legislator i, which is constant across all T decisions, and let $\mathbf{x} \in \mathbb{R}^{2N}$ be the matrix of ideal policies. Let $p^t \in \{(0,0), \rho^t\}$ be the outcome on issue t, let $\mathbf{p} = (p^1, ..., p^t, ..., p^T)$ be the sequence of outcomes and similarly, let $\boldsymbol{\rho} = (\rho^1, ..., \rho^t, ..., \rho^T)$ be the sequence of proposals. Let $\|\cdot\|$ be the Euclidean norm, then for any $i \in \mathcal{N}$, let $v_i(p^t) = -\left\|p^t - x^i\right\|$ denote the utility of i from each of the t choices made by the assembly, and let $u_i : \mathbb{R}^{2T} \longrightarrow \mathbb{R}$ denote the utility of agent i from the whole game, so $u_i(\mathbf{p}) = \sum_{t=1}^{T} v_i(p^t) = \sum_{t=1}^{T} -\left\|p^t - x^i\right\|$.

The distribution of ideal points is symmetric with respect to both axes. Formally, construct a $2(K+1) \times 2(K+1)$ grid for some finite $K \in \mathbb{N}$ such that for any $m, n \in \{-K, -K+1, ..., K-1, K\}$, the number of agents with ideal policy (m,n) is $N_{m,n}$ and $\sum_{m=-K}^{K} \sum_{n=-K}^{K} N_{m,n} = N$. The symmetry assumption is that $N_{m,n} = N_{-m,n} = N_{m,-n} =$

$N_{-m,-n}$. Further, I assume that $N_{m,n} \geq 1$ for any $m, n \in \{-1, 0, 1\}$ so that there are at least nine agents in the assembly, and that $N_{0,0} < 2\sum_{k=1}^{K} N_{k,k}$ so that there are not too many agents with an ideal policy at the origin.

Note that given any distribution of finitely many ideal points that are rational numbers, we can always construct a grid so that the ideal points lie in this grid, hence locating the ideal points in a grid only restricts them to be rational numbers, and without loss of generality we restrict attention to ideal points that are integers in each dimension for the remainder of the paper. Except for the nine agents nearest to the origin of coordinates, all other positions of the grid might be empty. A sufficiently fine grid approximates arbitrarily close any continuous distribution of preferences that is symmetric with respect to both axes, including as examples a uniform distribution, a bivariate normal, or less standard shapes such as a sum of bivariate normals with two modes.

The distribution of ideal policies satisfies the radial symmetry condition detailed by Plott (1967), by which for any given agent with an ideal policy in some direction away from the status quo, there is another agent with an ideal policy in the exact opposite direction. At the end of the paper I show that the results extend to more general distributions of preferences that do not satisfy radial symmetry. As I explain below, I make this restrictive assumption on the distribution of preferences deliberately, to study a harder case in which existing theories of party formation do not predict the formation of parties. With preferences that satisfy radial symmetry, the status quo policy $(0, 0)$ is a median in all directions and a Condorcet winner, that is, the status quo defeats any other policy in pairwise comparisons. Therefore, if agents vote their true preference in the assembly, every proposal fails.

I assume that, at a cost, agents can make binding commitments to coordinate their votes on all decisions. The timing depends on whether the agenda is exogenous or endogenous. I first consider an exogenous agenda, in propositions 1 and 2. I then consider an endogenous agenda, leading to propositions 3 and 4. If the agenda is endogenous, a given subset of $\mathcal{N}$ are potential agenda setters; this subset is fixed and common knowledge. The timing for either endogenous or exogenous agendas is as follows:

First, if the agenda is endogenous, Nature chooses one legislator among the set of potential agenda setters. The chosen agenda setter proposes an agenda $\boldsymbol{\rho} \in \mathbb{R}^{2T}$, which specifies a proposal $\rho^t \in \mathbb{R}^2$ for each decision t. The proposed agenda is revealed and becomes common knowledge. If the agenda is exogenous, this step is skipped, and agents know that Nature chooses ρ^t at a later stage.

Second, every agent can issue an invitation to any subset of other agents to form a voting bloc that includes the proposer. These invitations all become common knowledge as well. An invitation takes the form $\mathcal{N}_i \subseteq \mathcal{N} \cup \emptyset$, where $\mathcal{N}_i$ is the set of agents invited to join i in a voting bloc, and necessarily $i \in \mathcal{N}_i$ if $\mathcal{N}_i$ is not the empty set.

Third, each agent who receives an invitation to form a voting bloc can accept at most one invitation, or she can reject them all. If every $j \in \mathcal{N}_i$ accepts invitation $\mathcal{N}_i$, then $\mathcal{N}_i$ forms a voting bloc and every $j \in \mathcal{N}_i$ bears a cost $c_i > 0$.

Fourth (only if the agenda is exogenous), Nature chooses $\boldsymbol{\rho}$ by drawing ρ^t independently for each t from a uniform distribution in $[-1, 1]^2$. The drawn $\boldsymbol{\rho}$ becomes public knowledge.[1]

Fifth, legislators who are members of a bloc meet on a caucus and they vote on each decision, choosing between ρ^t or the status quo. Every voting bloc $\mathcal{N}_i$ coordinates by simple

[1]The results are robust for any distribution in $\mathbb{R}^2$ that is symmetric with respect to each axis and has positive density around the origin. The proofs are available in the online supplementary material.

majority: For any voting bloc $\mathcal{N}_i$ that forms at stage three, and any decision t, if a simple majority of $\mathcal{N}_i$ votes in favor of ρ^t in the caucus, every $j \in \mathcal{N}_i$ votes in favor on this issue in the assembly; if a simple majority votes against the proposal in the caucus, they all vote against the proposal in the assembly; and if they tie in the caucus, agents are free to vote as they wish in the assembly.

Sixth, the assembly meets and votes sequentially on each proposal ρ^t, deciding by simple majority. Independent agents vote as they wish, while members of a bloc are bound by their commitment to follow the outcome of the caucus of their bloc.

As noted, agents can make binding commitments to coordinate their votes within a bloc, and vote together in the assembly. The cost of organizing a voting bloc captures the difficulty of making these commitments. If it is possible to punish defectors ex-post at no cost, if only by social sanctions such as excluding them from a relevant social network, or if there exist bonds or deposits that legislators can put up front as guarantee that they will not defect from the bloc, then these commitment technologies suffice to enforce the coordination of votes. Alternatively, we can assume that the cost of organizing a voting bloc includes the cost of hiring external agents to act as enforcers and punish defectors. In any case, the assumption of binding commitments is made for simplicity, and I discuss at the end of paper how the main results on the formation of parties are robust if commitments are not feasible.

The strategy of each agent consists of at least three elements: The decision to issue invitations to form a voting bloc, the decision to accept one of these invitations, and the vote on each of the issues.[2] If the agenda is endogenous, the strategy of the potential agenda

[2] The protocol to form a voting bloc is similar to Hart and Kurz's (1983) coalition game Γ, first introduced

setters has an additional element: The agenda they choose. Furthermore, if the agenda is endogenous, the decisions to form voting blocs are a function of the chosen agenda.

The solution concept I use is Subgame Perfect Nash Equilibrium in iterated weakly undominated strategies.

Note that at the voting stages, once voting blocs have formed, only sincere voting survives the iterated elimination of weakly dominated strategies. Sincere voting on decision T is weakly dominant on the last subgame. On decision $t' < T$, if on every decision $t > t'$ agents vote sincerely, then by backward induction it is weakly dominant to vote sincerely on decision t as well. Sincere voting, for members of a bloc, means voting their preference in the caucus. In the assembly, they do not make a strategic decision; rather, they are bound to follow the dictates of their bloc. Given that only sincere voting survives the iterative elimination of weakly dominated strategies, I assume that agents correctly anticipate sincere voting on the part of every other agent at all stages and all subgames, and I consider a reduced strategy space that deals only with the agenda and the decisions about forming voting blocs. I rule out abstention, assuming that agents who are indifferent (a non-generic event) vote in favor of the proposal.

The first result is a partial equilibrium result, solving the game in which only one agent has the ability to invite others to form a voting bloc. Let agent l have ideal policy $x^l = (x^l_1, x^l_2)$ such that $x^l_1 \neq 0$ and $x^l_2 = 0$. Recall that c_i is the cost of joining a group founded by agent i.

by von Neumann and Morgenstern (1944) –for an overview of the coalition formation literature, see Ray (2007). Since all the legislators in a voting bloc must agree to join in order for the bloc to form, it must be that the formation of a voting bloc benefits every member of the party.

Proposition 1 *Assume the agenda is exogenous and* $c_i = \infty$ *for any* $i \in \mathcal{N} \backslash \{l\}$. *An equilibrium exists, and there exists a threshold* $\bar{c} > 0$ *such that if* $c_l < \bar{c}$, *then in every equilibrium a voting bloc forms and in each decision* t *with positive probability the assembly chooses proposal* ρ^t *over the status quo.*[3]

The literature on the endogenous formation of parties in a legislative assembly has noted that parties form to distribute pork (Baron 1989 and Jackson and Moselle 2002), to control the agenda (Cox and McCubbins 1993 and 2007, Diermeier and Vlaicu 2008) or to eradicate cycles[4] and to solve the instability inherent to political competition in multiple dimensions (Aldrich 1995). I show that legislators have incentives to coordinate their votes, coalescing into a voting bloc purely for ideological gain, even if they have no control over the agenda, and even in the absence of majority cycles or instability. In proposition 1 I show that a set of agents who coordinate their votes forming a voting bloc succeeds in defeating a Condorcet winning status quo policy.

If the status quo policy is a Condorcet winner, standard theories of policy-making predict that the status quo will be the policy outcome. In Krehbiel's (1998) *pivotal politics* theory, the Condorcet winner (in one dimension, the median ideal policy) lies inside the *gridlock* area, where policies cannot be changed. Normative reasons as well indicate that a Condorcet winner status quo policy should not be changed: Any change benefits only a minority of agents, and is detrimental for a majority. If utilities are linear or concave in distance to

[3] The probability is ex-ante. Ex-post, once a generic proposal ρ^t is revealed, it is either approved with certainty or rejected with certainty.

[4] A majority cycle occurs when a simple majority of voters strictly prefers an alternative a to b, a different simple majority strictly prefers b to c, and yet another simple majority of voters strictly prefers c to a.

the ideal policy, any deviation from the Condorcet-winning policy generates a loss in social welfare. Nevertheless, a group of legislators who share a common interest in one dimension of policy, but diverge in another dimension, can coalesce to coordinate their votes and win a majority to defeat the Condorcet winner and move the policy away from the status quo.

Note that members of a bloc benefit in expectation. Ex post, if the policy outcome moves away from the status quo, it generates an aggregate gain for the bloc, but the change can make some members worse off. If the number of decisions is large, it is more likely that every member benefits ex post as well. The ex ante benefit occurs even if there is only one decision.

The radial symmetry condition on the distribution of preferences does not drive the result. On the contrary, I impose the condition to stack the deck against the formation of a party, and to distinguish my argument from Aldrich's (1995) interpretation of parties as means to avoid instability. I show that even though there are no majority cycles to exploit, a voting bloc still manages to attain a net gain in expected utility by changing the policy outcome. I prove that the result is robust to perturbations on preferences that destroy radial symmetry at the end of the section.

The formation of a single voting bloc is not an equilibrium of the complete game in which any agent can invite others to form a voting bloc. If agents receive more than one invitation to join a bloc, coordination issues arise. For instance, if legislators i and i' both invite legislators j and j' to form a three person voting bloc, a bloc forms if j and j' coordinate to accept the same invitation, but it fails to form otherwise. If legislators j and j' would benefit from forming either bloc but they fail to do so because they accept different invitations, they are in a coordination failure.

Definition 1 *Given the strategy of every $i \notin A$, the strategy profile of a set of agents A is a coordination failure if*

(i) No $i \in A$ joins any voting bloc and

(ii) Every $i \in A$ would be strictly better of in expectation if A formed a voting bloc.

The definition of a coordination failure is contingent on the strategy profile of the other agents, so the strategies of a set of agents are a coordination failure only given what other agents do. The expectation of utilities is with respect to the stochastic realization of the agenda if it is exogenous, and to the realization of mixed strategies by other agents. Note that this definition of coordination failure is very narrow. It excludes coordination failures with agents who join another voting bloc, even if these agent would prefer to leave their blocs and form a different bloc. The definition only applies to cases that we may deem as complete failures, where agents who would all benefit from forming a voting bloc, all end up being independent. Presumably, agents should be able to avoid these coordination failures. If so, in equilibrium, at least two voting blocs form.

Proposition 2 *Assume the agenda is exogenous and $c_i = c > 0$ for any $i \in \mathcal{N}$. There is no one-bloc equilibrium without coordination failures. Furthermore, there exists $\bar{c} > 0$ such that if $c < \bar{c}$, then there is an equilibrium with two voting blocs and no coordination failures on the equilibrium path, and in any equilibrium without coordination failures along the equilibrium path, at least two voting blocs form.*

In the fully symmetric environment that I have described, a single voting bloc cannot gain an advantage, because an opposing set of legislators is able to form its own bloc to thwart any gain. If the cost of forming a party is low enough, by proposition 1, there is a set of agents who benefit from forming a voting bloc; if no bloc forms, these agents are in a coordination failure. The proof of existence of an equilibrium with two blocs is constructive. In the equilibrium I construct, agents separate into blocs according to their preference in one dimension: agents to the left of the vertical axis join a bloc, agents to the right join another bloc, and agents on the vertical axis split between either bloc or remaining independent. Coordination failures do not occur on the equilibrium path, but might occur off the equilibrium path. Both blocs are of size less than minimal winning.

I next consider an assembly where some members have a built in advantage, a position of privilege or power. Assume the agenda is endogenous, and there is a unique agenda setter, who has positive agenda power: She makes policy proposals that are put to a vote without amendments. The timing of the game must be different with an endogenous agenda, because it is unrealistic that the agenda setter can persuade other agents to join a bloc without explaining the proposals she plans to put to a vote. Accordingly, the agenda setter announces her agenda at the first stage, before agents send out and then accept invitations to form parties.

Assume that any agent can invite others to form a bloc at a cost c to each of its members, but the agenda setter a enjoys a technological advantage to coordinate, so she can invite other agents to form a bloc at a cost c_a to its members, with $c_a \leq c$. Perhaps the agenda setter a has a position of power within or outside the assembly that enables her to offer carrots and rewards outside the model to induce other agents to coordinate, while these tools to foster discipline and coordination are not available to other agents.

If the cost of coordination is lower for the agenda setter, she is able to exploit her privilege. She forms a voting bloc and introduces an agenda of proposals $\boldsymbol{\rho}$ that passes and benefits every member of her bloc. Agents with the opposite preferences could render this bloc ineffective by forming their own bloc, but the agenda setter prevents this by letting $\boldsymbol{\rho}$ be sufficiently close enough to the status quo so that the opposition does not have enough incentives to overcome the higher costs it faces when it forms its own voting bloc.

Proposition 3 *Assume $T \geq 2$, the agenda is endogenous, agent a with ideal policy $x^a \neq (0,0)$ is the agenda setter and $c_i = c > 0$ for any $i \in \mathcal{N} \backslash \{a\}$. There exists $\bar{c} > 0$ such that if $c_a < \bar{c}$, in any equilibrium with positive probability at least one voting bloc forms and the assembly chooses an outcome different from the status quo.*

The agenda setter can use her advantage to form a unique voting bloc by proposing an agenda that moves the policy away from the status quo, but keeps it close enough so that the losses for other agents are not sufficient to motivate them to form a second bloc. The agenda setter is constrained only to the extent that other agents can coalesce cheaply: An opposition that can easily coordinate would not let a unique voting bloc form unless the agenda is close to the status quo, while an opposition that faces great difficulties in coordinating is faced with larger policy deviations towards the preference of the agenda setter and her bloc.

Since the space of possible agendas is infinite, existence of equilibrium may become an issue. A shortcut to prove existence is to turn the game into a finite one by assuming that there are only finitely many feasible proposals on each decision so that the set of possible agendas is finite. A possible interpretation is that implemented policies can only change by discrete increments along each dimension, so proposals must lie on a grid, though this grid could be arbitrarily fine to approximate the continuous case.

Suppose the agenda setter is the agent with ideal preference $(1, 0)$ depicted in figure 1, and she proposes the following agenda: (ρ_1, ρ_2^{odd}) along the indifference curve of the agents with ideal policy $(-1, 1)$ for any t odd and $(\rho_1, -\rho_2^{odd})$, which lies along the indifference curve of the $(-1, -1)$ agents, for any t even, with $\rho_1, \rho_2^{odd} > 0$, and choosing (ρ_1, ρ_2^{odd}) just close enough to $(0, 0)$ so that no other agents have an incentive to propose a second voting bloc. That is, the agenda setter a proposes the most favorable pair of points in the shaded areas in figure 1 that are symmetric with respect to the horizontal axis and that do not lead the opposition to form a second bloc. In every decision, the agenda setter proposes to move the policy to the right. In half the decisions it proposes to move right and up; in the other half she proposes right and down. Members of the bloc benefit because they manage to trade votes in favor of right-and-up proposals by the right-and-down agents for votes in favor of right-and-down proposals by right-and-up agents. Members of the bloc extract a benefit because legislators with ideal policies to the left of the status quo fail to coordinate in a similar manner to prevent the passage of all these policies.

If no agent has an advantage in organizing voting blocs, I show that no single bloc can move the outcome away from the status quo. For any agenda and any voting bloc such that all its members benefit from the bloc, an opposing voting bloc consisting of the legislators with the exact opposite preferences can also form to prevent the passage of the policy proposals. But if the actions of two blocs cancel each other out, and the agenda is endogenous, the agenda setter is better off proposing the status quo policy in each decision, so that there is no political competition and no need to incur the costs of forming voting blocs to fight political battles that are not going to succeed. Nevertheless, we often observe two political parties in a legislature competing against each other, with at least one of the two parties pursuing an aggressive agenda that is eventually defeated in the assembly –for instance, in 2007 the Democrats in the US Senate put to a vote no less than a dozen failed legislative initiatives for troop withdrawals in Iraq.[5] Parties who engage in these fights incur the cost of coordinating members and trying to marshal their votes, while achieving no benefit in terms of change in policy outcomes.

I explain this aggressive agenda introducing uncertainty. I maintain the assumption that legislators are outcome oriented and care only for the implemented policy, but I relax the assumption of perfect commitment. Instead, I allow voting blocs to form, but I assume that with some exogenous probability λ, coordination fails. The cost of accepting an invitation to coordinate in a bloc is now not a sure investment but a risky one: With probability λ, the voting bloc fails, commitments are not binding, and the sunk cost of forming a bloc is wasted. With probability $1-\lambda$, the voting bloc works and enforces commitments. These probabilities are exogenous and independent across blocs. With uncertainty, an agenda setter has an incentive to propose an aggressive agenda in the hope that her bloc succeeds in coordinating, and the opposition does not.

Proposition 4 *Assume* $T \geq 2$, *the agenda is endogenous, the set of potential agenda setters is* $\{i \in \mathcal{N} : x^i = (x_1^i, 0), x_1^i \neq 0\}$, *the cost of coordination is* $c_i = c > 0$ *for any* $i \in \mathcal{N}$ *and there is uncertainty* $\lambda \in (0, 1/2)$ *about the enforcement of commitments. There exists* $\bar{c} > 0$ *such that if* $c < \bar{c}$, *in any equilibrium without coordination failures the agenda setter proposes an agenda* $\boldsymbol{\rho}$ *different from the status quo and at least two voting blocs form, and the with positive probability the assembly chooses a policy outcome away from the status quo.*

[5] http://www.senate.gov/legislative/LIS/roll_call_lists/vote_menu_110_1.htm

The chosen agenda setter puts forward an aggressive agendas that, if approved, changes the policy outcome away from the status quo and toward her policy preference in the hope that the bloc(s) who favor the proposal succeed(s) in coordinating while the other bloc(s) fail(s). In this case, the assembly chooses an outcome away from the status quo. Note that it is possible to endogenize the agenda setter as well, and not just the agenda. Instead of letting nature choose an agenda setter from an exogenously given set of potential agenda setters, let the first stage be modified to be one of endogenous candidacy for the agenda setting position, followed by voting, as in the citizen candidate model of Besley and Coate (1997), with the crucial difference that the winner becomes an agenda setter, instead of becoming a policy-maker. Then, in a two candidate equilibrium with agents a and b with ideal policies $(x_1^a, 0)$ and $(x_1^b, 0)$ such that $x_1^a = -x_1^b$, both candidates tie, one is chosen randomly, and just as in proposition 4, at least two blocs form.

More general preferences: Relaxing radial symmetry

I assumed that preferences lie on a grid and satisfy radial symmetry so that the status quo is a Condorcet winner to show that voting blocs form and change the outcome even in the absence of cycles or intransitivity of majority preferences, thus distinguishing my argument from the explanations of party formation by Aldrich (1995). Aldrich argues that parties form to prevent majority preferences from cycling and to achieve a stable political outcome. I show that parties form to move the policy outcome in their preferred direction, away from the socially optimal, Condorcet winning status quo.

If preferences do not satisfy radial symmetry, there is no Condorcet winning outcome and there is a set of outcomes that are preferred by a majority of agents over the status quo. However, each of these outcomes is itself majority-preferred by some other. With such a distribution of preferences, Cox (1987) predicts and Bianco and Sened (2005) find that the policy outcome lies somewhere in the uncovered set McKelvey (1986). An alternative x *covers* y if x beats y and any alternative z that beats x also beats y according to majority preferences. The alternatives that are not covered constitute the uncovered set. If r is the radius of the smallest ball B that intersects all the median hyperplanes, then the uncovered set is contained within a ball of radius $4r$ centered at the center of B.

With more general preferences that do not satisfy radial symmetry, my theory shows that voting blocs form and move the outcome outside the uncovered set.

Perturb the preferences as follows. Given the original profile of ideal points $\mathbf{x}$ which satisfies radial symmetry, for each $i \in \mathcal{N}$, let $\widetilde{x}^i$ be the new ideal point of the agent, such that $\widetilde{x}^i \in N(x^i, \varepsilon)$, where $N(x^i, \varepsilon)$ is the neighborhood of size ε around x^i. Let $\tilde{\mathbf{x}} \in \mathbb{R}^{2N}$ denote the profile of ideal points of every agent, which along with the assumption of Euclidean preferences, determines the preferences of every agent on each single decision. This new preference profile is more general, since it relaxes radial symmetry. If the perturbed preferences are close enough to the original preferences on a grid, propositions 1-4 hold, subject to the appropriate restatement.

Proposition 5 *Let the preferences profile be given by* $\tilde{\mathbf{x}}$.

1) Assume the agenda is exogenous and $c_i = \infty$ *for any* $i \in \mathcal{N} \backslash \{l\}$, *where* $x^l = (x_1^l, 0)$ *and* $x_1^l \neq 0$. *There exist* $\overline{\varepsilon} > 0$ *and* $\bar{c} > 0$ *such that for any* $\varepsilon \leq \overline{\varepsilon}$, *if* $c_l < \bar{c}$ *in every equilibrium a voting bloc forms and with positive probability, the assembly chooses an outcome outside the uncovered set.*

2) Assume that the agenda is exogenous and $c_i = c > 0$ *for any* $i \in \mathcal{N}$. *There exist* $\bar{\varepsilon} > 0$ *and* $\bar{c} > 0$ *such that for any* $\varepsilon \leq \bar{\varepsilon}$, *if* $c < \bar{c}$ *there is an equilibrium with two voting blocs and no coordination failures along the equilibrium path, and there is no equilibrium without coordination failures along the equilibrium path in which less than two voting blocs form.*

3) Assume $T \geq 2$, *the agenda is endogenous, agent* a *with ideal policy* $\widetilde{x}^a \neq (0,0)$ *is the agenda setter and* $c_i = c > 0$ *for any* $i \in \mathcal{N} \backslash \{a\}$. *There exist* $\bar{\varepsilon} > 0$ *and* $\bar{c} > 0$ *such that for any* $\varepsilon \leq \bar{\varepsilon}$, *if* $c_a < \bar{c}$, *then with positive probability at least one voting bloc forms and the outcome in at least one decision is outside the uncovered set.*

4) Assume $T \geq 2$, *the agenda is endogenous, the set of potential agenda setters is* $\{i \in \mathcal{N} : x^i = (x_1^i, 0), x_1^i \neq 0\}$, *the cost of coordination is* $c_i = c > 0$ *for any* $i \in \mathcal{N}$ *and there is uncertainty* $\lambda \in (0, 1/2)$ *about the enforcement of commitments. There exist* $\bar{\varepsilon} > 0$ *and* $\bar{c} > 0$ *such that for any* $\varepsilon \leq \bar{\varepsilon}$, *if* $c < \bar{c}$ *in any equilibrium without coordination failures, at least two voting blocs form.*

Note that the restatements of propositions 1-4 amount only to eliminate references to the Condorcet winner, which no longer exists, and to note that the outcome is not only away from the status quo, but outside the uncovered set.

In summary, agents coalesce to coordinate their votes regardless of the existence or inexistence of Condorcet winners or cycles in the majority preference, and regardless of whether the agenda is exogenous or endogenous. If one agent has an advantage in the formation of parties, a unique bloc that benefits this agent forms. Otherwise, at least two blocs form.

Discussion

I have presented a theory of coalition formation in which members of an assembly who have spatial preferences on a multidimensional choice set coordinate to form voting blocs. These voting blocs function as disciplined parties so that all members cast their votes together in the assembly.

I have shown that if one agent has an exogenous advantage that allows her to coordinate a voting bloc at a lower cost, in equilibrium one party forms around this agent, and the policy outcome moves away from the status quo, even if this status quo is both the Condorcet winning policy and the social welfare maximizing policy.

On the other hand, if no individual agent has a coordination advantage, in any equilibrium at least two parties form.

These findings are robust to the consideration of either an exogenous random agenda or an endogenous strategic agenda, and they are also robust to perturbations of the distribution of policy preferences so that majority cycles occur and there is no Condorcet winning policy.

The results are also robust if we relax the assumption of commitment, and study blocs that cannot enforce discipline. In this case, the outcome of the internal meeting of the bloc is only a voting prescription that the members of the party can follow or ignore in the assembly, without more than infinitesimal punishments for ignoring it. If there is no utility loss from ignoring the prescription, following it against sincere preferences is a weakly dominated strategy. However, if a member who deviates from the voting prescription of her bloc incurs an arbitrarily small cost $\varepsilon > 0$, even with undominated strategies coordination can affect the outcome in any event in which the voting outcome in the assembly is decided by more than one vote, so no individual agent is pivotal and no agent has an individual incentive to deviate from her bloc's voting prescription (see Eguia 2010).

Blocs can avoid the violation of their voting prescriptions by modifying the simple majority internal rule to stipulate that agents are free to vote as they wish whenever the coordination of votes by all parties would result in an outcome decided by one vote. With these internal rules, members have no incentives to deviate from their bloc's voting prescription in all other decisions. Without commitment, blocs become coordination devices for decisions where no agent is individually pivotal, and they still affect some outcomes and benefit their members in expectation, so agents have almost the same incentives to form blocs as in the case with binding commitment.

Appendix

Proposition 1

Proof. The game is finite, so existence follows directly from Nash's (1950) theorem.

For any x_1, x_2, let i_{x_1,x_2} denote an arbitrary agent i with $x^i = (x_1^i, x_2^i)$. Without loss of generality, let $x_1^l \geq 1$. Let $A_1 = \{i_{x_1,x_2} : x_1 > 0 \text{ and } x_1 - 1 \leq |x_2| \leq x_1\}$. Consider $\mathcal{N}_l = A_1 \cup \{l\}$. On each decision t, the bloc $\mathcal{N}_l$ favors proposal ρ^t if and only if $i_{1,0}$ favors it. If $i_{1,0}$ favors it, l favors it and either all $i_{x_1,x_2} \in A_1$ with $x_2 \geq 1$ or all $i_{x_1,x_2} \in A_1$ with $x_2 \leq -1$ favor it as well, constituting a majority of the bloc in favor. If $i_{1,0}$ prefers the status quo, either all $i_{x_1,x_2} \in A_1$ with $x_2 \geq 1$ or all $i_{x_1,x_2} \in A_1$ with $x_2 \leq -1$ prefer the status quo as well. Given that $\mathcal{N}_l$ never votes as a bloc against the preference of $i_{1,0}$, no ρ^t gathers a majority in the assembly if $i_{1,0}$ opposes it. Given a policy ρ^t, if $i_{1,0}$ and $i_{-1,1}$

and i_{x_1,x_2} favor it for all x_1, x_2 such that $x_1 \geq 2$ and $x_2 = -x_1 + 2$, if bloc $\mathcal{N}_l$ forms then ρ^t passes in the assembly. Similarly, ρ^t passes if $i_{1,0}$ and $i_{-1,-1}$ and i_{x_1,x_2} favor it for all x_1, x_2 such that $x_1 \geq 2$ and $x_2 = x_1 - 2$. Since the slopes of the indifference curves of $i_{1,0}$ and i_{x_1,x_2} such that $x_1 \geq 2$ and $x_2 = -x_1 + 2$ at $(0,0)$ are all greater than 1 and the indifference curve of $i_{-1,1}$ at $(0,0)$ is exactly 1, the set of proposals that all these agents favor has a non empty interior. Given the symmetry of the distribution of preferences with respect to the horizontal axis, the area of proposals that pass if $\mathcal{N}_l$ forms is divided into two areas, symmetric with respect to the horizontal axis.[6]

I want to show that for any pair of proposals $\rho^k = (y_1, y_2)$ and $\rho^{k\prime} = (y_1, -y_2)$, where $y_1, y_2 \in [0,1]$ are such that

$$v_{i_{1,0}}((y_1, y_2)) + v_{i_{1,0}}((y_1, -y_2)) \geq 2v_{i_{1,0}}((0,0)), \tag{1}$$

it follows

$$v_i((y_1, y_2)) + v_i((y_1, -y_2)) \geq 2v_i((0,0))$$

for every $i \in \mathcal{N}_l$. Let $z_1 \leq y_1$ be such that

$$v_{i_{1,0}}((z_1, y_2)) + v_{i_{1,0}}((z_1, -y_2)) = 2v_{i_{1,0}}((0,0)). \tag{2}$$

Since

$$v_i((z_1, y_2)) + v_i((z_1, -y_2)) \leq v_i((y_1, y_2)) + v_i((y_1, -y_2))$$

for any $i \in \mathcal{N}_l$, to establish inequality 1 it suffices to show that

$$v_i((z_1, y_2)) + v_i((z_1, -y_2)) \geq 2v_i((0,0))$$

[6] I illustrate the two sets of policies that pass, shaded, in figure 1 for an assembly with $K = 1$ and $N_{m,n} = 1$ for $m, n = \{-1, 0, 1\}$, so $N = 9$, and with $x^l = (1,0)$. In this case, ρ^t passes if and only if $i_{1,0}$ and either $i_{-1,1}$ or $i_{-1,-1}$ favors it.

for every $i \in \mathcal{N}_l$. Note that inequality 2 implies

$$\begin{aligned} (z_1 - 1)^2 + y_2^2 &= 1 \\ z_1 &= 1 - \sqrt{1 - y_2^2}. \end{aligned}$$

I want to show that for any (x_1, x_2) such that $x_1 \geq 1$ and $-x_1 \leq x_2 \leq x_1$, and any (z_1, y_2) such that $z_1 = 1 - \sqrt{1 - y_2^2}$ and such that $z_1, y_2 \in [0, 1]$,

$$[(x_1 - z_1)^2 + (x_2 - y_2)^2]^{1/2} + [(x_1 - z_1)^2 + (x_2 + y_2)^2]^{1/2} \leq 2[x_1^2 + x_2^2]^{1/2}.$$

Algebraic manipulations (omitted here, but available online in the supplementary material) of this expression yield

$$2(x_1^2 + x_2^2) + 2(x_1^2 + x_2^2)(x_1 - 1)\sqrt{1 - b^2} \leq b^2 x_2^2 + 2x_2^3 + 2x_1 x_2^2$$

Since $0 \leq y_2 \leq 1$ and $x_1 \geq 1$, it suffices to check

$$2(x_1^2 + x_2^2)x_1 \leq b^2 x_1^2 + 2(x_1^2 + x_2^2)x_1$$

which holds for any x_2. Furthermore, if $y_2 \neq 0$ the left hand side is strictly less than $2(x_1^2 + x_2^2)x_1$ and the last expression holds with strict inequality.

Given any point (y_1, y_2) that passes if $\mathcal{N}_l$ forms, $(y_1, -y_2)$ passes as well. Given the uniform density function of the exogenous agenda, the density function of proposals (y_1, y_2) and $(y_1, -y_2)$ is the same. It follows that every $i \in \mathcal{N}_l$ strictly benefits from the coordination of votes inside $\mathcal{N}_l$. Let $\bar{c} > 0$ be the minimum gain in individual expected utility among all agents in $\mathcal{N}_l$. If $c_l < \bar{c}$, any strategy in which any $i \in \mathcal{N}_l$ rejects the invitation to form $\mathcal{N}_l$ is weakly dominated. Then, invitation $\mathcal{N}_l = A_1 \cup \{l\}$ makes agent l strictly better off than invitation $\mathcal{N}_l' = \emptyset$.

Consider the subgame after l makes any arbitrary invitation. These subgames are finite, hence they have a Nash equilibrium. Select the subgame and Nash equilibrium that generates the highest payoff to l. If $c < \bar{c}$ defined above, we have shown that the subgame with $\mathcal{N}_l = A_1 \cup \{l\}$ yields a payoff strictly greater than the benchmark with no blocs. Thus the highest payoff among all subgames is higher than the payoff with no blocs, which necessarily implies that the proposals defeat the status quo with positive probability, which occurs only if a voting bloc forms. It follows that the equilibrium strategy of agent l must be to issue an invitation to form a voting bloc; in equilibrium this invitation is accepted, and the policy proposal defeats the status quo at each stage with positive probability. ■

Proposition 2

Proof. I first rule out equilibria without coordination failures and with no blocs. As shown in the proof of proposition 1, $\exists A \subset \mathcal{N}$ and $l \in A$ such that if $c_l > 0$ is low enough, every $i \in A$ is strictly better off if A forms a unique voting bloc. Hence if $c > 0$ is low enough, an outcome with no blocs is a coordination failure by A.

I next rule out equilibria without coordination failures and with one bloc. Consider a strategy profile such that the set of agents A forms a voting bloc and no other bloc forms. In order for every $i \in A$ to be best responding, it must be that each i weakly benefits from the formation of A. Pair all agents with $x^i \neq (0,0)$ as follows: For any i with $x^i = (x_1^i, x_2^i)$, let $j(i) : \mathcal{N} \longrightarrow \mathcal{N}$ be a one-to-one mapping such that $x^{j(i)} = (-x_1^i, -x_2^i)$. For any i and $j(i)$ and any p^t, if

$$v_i(p^t) - v_i((0,0)) \geq 0 \Longrightarrow v_{j(i)}((0,0)) - v_{j(i)}(p^t) \geq v_i(p^t) - v_i((0,0))$$

and this inequality is strict for any p^t that is not a convex combination of x^i and $x^{j(i)}$; that is, the inequality is strict for a generic p^t. Let $B = \{j(i) : i \in A\}$. If A and B each form a bloc and no other agent forms another bloc, they cancel each other out and no proposal passes. Hence if the agents in A are weakly better off forming A, then the set of agents in B are strictly better off forming B in response, and the formation of a unique bloc by A is a coordination failure for B.

Third, I show existence of an equilibrium with two blocs. Let $A_1 = \{i : x^i = (x_1^i, x_2^i)$ with $x_1^i \geq 1\}$, let $B_1 = \{i : x^i = (x_1^i, x_2^i)$ with $x_1^i \leq -1\}$, and let $C_1 = \{i : x^i = (x_1^i, x_2^i)$ with $x_1^i = 0\}$. Recall i_{x_1,x_2} denotes an agent with ideal policy (x_1, x_2). At the invitation stage, $\mathcal{N}_{i'} = \mathcal{N}_{i''} = A_1$ for arbitrary agents $i', i'' \in A_1$; $\mathcal{N}_{j'} = \mathcal{N}_{j''} = B_1$ for arbitrary agents $j', j'' \in B_1$; and $\mathcal{N}_{k'} = \mathcal{N}_{k''} = C_1$ for arbitrary agents $k', k'' \in C_1$ and $\mathcal{N}_i = \emptyset$ for any other agent. At the acceptance stage, given these six proposals, all $i \in A_1$ including i'' accept the invitation $\mathcal{N}_{i'}$, all $j \in B_1$ including j'' accept the invitation $\mathcal{N}_{j'}$ and all members of C_1 reject the invitations to form C_1. If i' and/or j' deviate at the invitation stage, agents accept the invitations by i'' and j'' instead. All agents in $A_1 \cup B_1$ ignore any other individual deviation at the invitation stage and continue to accept the invitations by i' and j'. Agents in C_1 ignore deviations and continue to reject the invitation to form C_1 and any other they may receive as long as ignoring them is not dominated. If they receive an invitation to form a bloc D instead and accepting this invitation weakly dominates rejection of all invitations, then all $i \in C_1$ except an arbitrary $k \in C_1 \cap D$ accept an invitation to join C_1, while k accepts the new invitation.

With these acceptance strategies, no agent has an incentive to individually deviate at the invitation stage, since the outcome does not change following the deviation. It remains to be shown that these acceptance strategies are best responses.

Given that the distribution of ideal policies inside A and B are symmetric with respect to the horizontal axis, the set of proposals that can pass if either A or B forms is symmetric with respect to the horizontal axis. If only A forms, only policies to the right of the origin pass; if only B forms, only policies to the left of the origin pass. Given that the set of proposals that can pass if only A forms is symmetric to the horizontal axis and completely to the right of the vertical axis, every $j \in B$ is hurt. Analogously, if only B forms, every $i \in A$ is hurt. If both A and B form, the outcome is $(0, 0)$. It follows, if c is small enough, that all members of A and B are better off accepting their invitations to form A and B. Given that A_1, B_1 form, the policy outcome is $(0, 0)$ regardless of whether C_1 forms or not, hence it is a best response not to form it.

Following a deviation at the proposal stage that proposes the formation of a voting bloc D with at least two members of $A_1 \cup B_1$, it is a mutual best response for all the invited members who belong to $A_1 \cup B_1$ to reject the invitation, since D is not going to form given that the other agent(s) reject(s) the invitation. Following a deviation that proposes the formation of D with only one member of $A_1 \cup B_1$, given that at most one member of C_1 accepts this invitation, D is not going to form, so the member of $A_1 \cup B_1$ is better off rejecting D to form A_1 and B_1 instead. For any member $i \in C_1$ invited to join D, since D is not going to form regardless of the action by i, rejecting the invitation to join D is a best response. If rejecting all invitations is a weakly dominated strategy, dominated by accepting D, for c low enough, accepting the invitation to join C_1 and accepting the invitation to join D are both undominated. Given that the member of $A_1 \cup B_1$ invited to D rejects D and prefers to have A_1 and B_1 form instead, any acceptance strategies such that C_1 does not form is a mutual best response for the members of C_1, in particular, agent k accepting D and all others accepting C_1 is a mutual undominated best response. Finally, any invitation to form a bloc D without any member $A_1 \cup B_1$ is rejected because given that blocs A_1 and B_1 form, a new bloc $D \subseteq C_1$ either has no effect on policy outcomes, or if it does, it must hurt at least one of its members ex ante, hence at least this member rejects its formation. Hence the described strategy profile is an undominated equilibrium. ■

Proposition 3

Proof. Without loss of generality, let $x^a = (x_1^a, x_2^a)$ be such that $x_1^a \geq 1$ and $|x_2^a| \leq x_1^a$. Suppose a proposes the following agenda of proposals $\boldsymbol{\rho}$: $\rho^t = (\rho_1^t, \rho_2^t)$ for every t odd except $t = T$ (if T is odd, let $\rho^T = (0,0)$), and $\rho^t = (\rho_1^t, -\rho_2^t)$ for every t even, where $\rho_1^t > 0$ and (ρ_1^t, ρ_2^t) is such that $v_{i_{-1,1}}((\rho_1^t, \rho_2^t)) = v_{i_{1,0}}((0,0))$ and $v_{i_{1,0}}((\rho_1^t, \rho_2^t)) > v_{i_{1,0}}((0,0))$. This is not uniquely defined, pick any agenda satisfying these conditions. Suppose a proposes the formation of the bloc $\mathcal{N}_a = \{i_{x_1,x_2} : x_1 \geq 1, | \, |x_2| \leq x_1\}$. If $\mathcal{N}_a$ forms, the assembly chooses ρ^t over the status quo at every decision t and, as shown in the calculations of proposition 1, $u_i(\boldsymbol{\rho}) > u_i((\mathbf{0}, \mathbf{0})^T)$ for every $i \in \mathcal{N}_a$. Given c, if $||(\rho_1^t, \rho_2^t)||$ is sufficiently small, the gain of forming any bloc is less than c and it is weakly dominated to accept any invitation to join a bloc that was not issued by the agenda setter a. Let $\bar{c} = \min_{\{i \in \mathcal{N}_a\}} u_i(\boldsymbol{\rho}) - u_i((0,0)^T) > 0$, so that $\bar{c}$ is equal to the benefit in terms of policy outcomes enjoyed by the member of the bloc who benefits the least if the bloc forms. If $c_a < \bar{c}$ and the agenda setter proposes $\mathcal{N}_a$, every $i \in \mathcal{N}_a$ strictly benefits if $\mathcal{N}_a$ forms. Since accepting an invitation by any other agent given agenda $\boldsymbol{\rho}$ is weakly dominated, not accepting invitation $\mathcal{N}_a$ is iteratively weakly dominated, so every $i \in \mathcal{N}_a$ accepts it, the policy outcome changes in every issue and every

$i \in \mathcal{N}_a$ becomes strictly better off.

Consider all the subgames that follow after a announces any other agenda. All these subgames are finite, hence they have an equilibrium. Take an equilibrium strategy profile in each subgame. The agenda setter's optimal strategy in the game, given these equilibrium strategy profiles for every subgame, is to choose the agenda that leads to the subgame with the highest payoff to the agenda setter. I showed that there is one agenda that leads to the formation of a voting bloc and makes the agenda setter strictly better off than the status quo. Hence, the agenda setter's best response at the initial stage must make her at least as well off than the agenda I have identified. To do so, it must change the policy outcome (at least with positive probability), and to change the policy outcome, at least one voting bloc must form. ■

Proposition 4

Proof. Without further loss of generality, let $x^a = (x^a_{1,}, 0)$ with $x^a_1 \geq 1$ be the chosen agenda setter. Let $A = \{i_{x_1,x_2} \in \mathcal{N} : x_1 \geq 1\}$.

Suppose a proposes agenda $\boldsymbol{\rho}$ such that $\rho^t = (0.1, 0.25)$ for every t odd except $t = T$ (if T is odd), $\rho^t = (0.1, -0.25)$ for every t even, and $\rho^T = (0, 0)$ if T is odd. Then $u_i(\boldsymbol{\rho}) > u_i((0, 0))$ for every $i \in A$. If $\boldsymbol{\rho}$ is the agenda and A forms a unique voting bloc, the assembly chooses ρ^t at every t. Let $b_{\min} = \min_{\{i \in A\}} u_i(\boldsymbol{\rho}) - u_i((0, 0))$. Given agenda $\boldsymbol{\rho}$, if $c < b_{\min}$ and no bloc forms, A is in a coordination failure. Let $\mathcal{N}_{\boldsymbol{\rho}} \subseteq \{\mathcal{N}_1, ..., \mathcal{N}_N\}$ be the set of voting blocs that form in a given equilibrium of the subgame after $\boldsymbol{\rho}$ becomes the agenda. Let $n = |\mathcal{N}_{\boldsymbol{\rho}}|$ be the number of blocs that form. As argued, if no blocs form, A is in a coordination failure, so in any equilibrium with no coordination failures, $n \geq 1$. Note that in the game with an endogenous agenda, $\boldsymbol{\rho}$ is common knowledge before invitations are sent out, so blocs form to try to implement $\boldsymbol{\rho}$ or in order to stop a rival bloc from doing so. Consider an arbitrary $\mathcal{N}_i \in \mathcal{N}_{\boldsymbol{\rho}}$. Let $\mathcal{N}_{-i} \equiv \mathcal{N}_{\boldsymbol{\rho}} \backslash \mathcal{N}_i$, that is, the subset of other blocs besides the one coordinated by agent i. In order for the members of $\mathcal{N}_i$ to be willing to form this bloc, it must be that, with some probability, the enforcement of commitments by $\mathcal{N}_i$ alters the outcome given the formation and probabilistic enforcement of commitments by the blocs in $\mathcal{N}_{-i}$. Therefore, given that the blocs in $\mathcal{N}_{-i}$ form, either at least some of the proposals in $\boldsymbol{\rho}$ pass with positive probability if $\mathcal{N}_i$ does not form, or they pass with positive probability if $\mathcal{N}_i$ forms (or both). Since by assumption each bloc with probability λ fails to enforce commitments even if it forms, it follows in either case that the policy proposals in $\boldsymbol{\rho}$ pass with positive probability if all the blocs in $\mathcal{N}_{\boldsymbol{\rho}}$ form. Let θ be this probability. Let $\theta_{\min}$ be the lowest value of θ over all the possible equilibria of the subgame after $\boldsymbol{\rho}$ is announced. Let $\beta = \theta_{\min}(u_a(\boldsymbol{\rho}) - u_a((0,0)^T))$. If $c < \beta$, in any equilibrium without coordination failures of the subgame after $\boldsymbol{\rho}$ becomes the agenda, the agenda setter a is in expectation strictly better off by $\beta - c$ than in the benchmark with the status quo outcome and no voting blocs.

Agent a chooses the agenda that maximizes the expected utility given the bloc proposal and acceptance strategies of all other agents. If $c < \beta$, agenda $\boldsymbol{\rho}$ described above generates a strictly positive net gain for a. If $\boldsymbol{\rho}$ is not the best response by a, the equilibrium agenda must yield a net gain to a greater than $\beta - c$. To do so, it must be an agenda away different from the status quo and at least one voting bloc must form (at least with positive probability).

Suppose only bloc $\mathcal{N}_i$ forms. Let $\hat{\boldsymbol{\rho}}$ be an equilibrium agenda. Then

$$(1-\lambda)(u_a(\hat{\boldsymbol{\rho}}) - u_a((0,0)^T)) \geq \beta - c$$
$$u_a(\hat{\boldsymbol{\rho}}) - u_a((0,0)^T \geq \frac{\beta - c}{1-\lambda}$$

so $\hat{\rho}^t$ must be bounded away from $(0,0)$ at least for some t. In particular, given $x^a = (x_1^a, 0)$ with $x_1^a \geq 1$, it must be that the expected $p_1^t > 0$, and bounded away from 0 so that p^t generates a utility gain to a of at least $\frac{\beta-c}{T}$ if it passes. Define $j(i)$ as in the proof of proposition 2. If only one voting bloc forms, it cannot contain a pair of agents i and $j(i)$ since they cannot both benefit from the formation of a bloc. Let $B = \{j(i) : i \in \mathcal{N}_i\}$. Consider any pair of agents (i, j) such that $i \in \mathcal{N}_i$ and $j = j(i)$.

If each proposal ρ^t is a convex combination of x^i and $x^{j(i)}$, then the proposals generate no aggregate gain or loss in utility for i and $j(i)$. But the expected outcome $E[p^t] > 0$ and it is bounded below, so the outcomes unambiguously benefit one agent more than the other, in particular, they benefit i and hurt $j(i)$, and this loss of utility to $j(i)$ is bounded below. Let d be this lower bound on the disutility that accrues to $j(i)$ if $\mathcal{N}_i$ forms a unique voting bloc. If B forms a voting bloc, with probability $(1-\lambda)^2$, B prevents this utility from materializing, by coordinating and cancelling out the effects of coordination in $\mathcal{N}_i$. Thus, if $c < (1-\lambda)^2 d$, $j(i)$ is better off if B forms; it follows that if B does not form, the agents in B are in a coordination failure.

If the proposals are not all convex combinations of x^i and x^j, then for any policy p^t that is not a convex combination of x^i and $x^{j(i)}$,

$$u_i(p^t) + u_{j(i)}(p^t) < u_i((0,0) + u_{j(i)}((0,0)).$$

and there is an aggregate loss of utility for i and $j(i)$. Since i must benefit from the formation of $\mathcal{N}_i$, it follows that $j(i)$ must be hurt. If c is low enough, by the same argument as above, $j(i)$ is better off if B forms a voting bloc, and if this bloc does not form, B is in a coordination failure. Hence, at least two voting blocs must form in any equilibrium without coordination failures. ■

Proposition 5

Proof. Follow the proofs of propositions 1-4, adapting them as follows.

1) Let $W(\mathcal{N}_1) \subset [-1, 1]^2$ be the set of proposals that beat $(0, 0)$ if $\mathcal{N}_1$ forms and $\varepsilon = 0$. Let $W^C(\mathcal{N}_1)$ be its complement. For any $\varepsilon > 0$, let $W_\varepsilon(\mathcal{N}_1)$ be the set of proposals that beat $(0, 0)$ for any preference profile $\tilde{\mathbf{x}}$ consistent with ε if $\mathcal{N}_1$ forms, and let $L_\varepsilon(\mathcal{N}_1)$ be the set of proposals that are defeated by $(0, 0)$ for any preference profile $\tilde{\mathbf{x}}$ consistent with ε if $\mathcal{N}_1$ forms. For any point $y = (y_1, y_2)$ and set V, let $d(y, V)$ be the Euclidean distance between y and V. For any $\delta > 0$ and for any $w \in W(\mathcal{N}_1)$ such that $d(w, W^C(\mathcal{N}_1)) \geq \delta$, $\exists \varepsilon > 0$ such that $w \in W_\varepsilon(\mathcal{N}_1)$. Similarly, for any $\delta > 0$ and any $z \in W^C(\mathcal{N}_1)$ such that $d(w, W^C(\mathcal{N}_1))$, $\exists \varepsilon > 0$ such that $w \in L_\varepsilon(\mathcal{N}_1)$, so as $\varepsilon \longrightarrow 0$, the set of policies that can pass given that $\mathcal{N}_1$ forms becomes arbitrarily close to the set of proposals that can pass given $\varepsilon = 0$. Since the density function of the exogenous agenda is uniform, it follows that as $\varepsilon \longrightarrow 0$ the correspondence of possible utilities if $\mathcal{N}_1$ forms given ε converges to the utility of forming $\mathcal{N}_1$ given $\varepsilon = 0$. If no blocs form, with $\varepsilon > 0$, the status quo is not always the outcome, but as $\varepsilon \longrightarrow 0$, the set of proposals that beat the status quo converges to the status quo itself, so the correspondence of possible utilities of not forming any bloc given ε converges to the utility of not forming any bloc given $\varepsilon = 0$. Hence, for any given c, as $\varepsilon \longrightarrow 0$, the net utility of forming $\mathcal{N}_1$ relative to the benchmark with no blocs converges

to the net utility of forming $\mathcal{N}_1$ given $\varepsilon = 0$. For $\varepsilon = 0$ and $c = 0$, members of $\mathcal{N}_1$ attain a strict net gain in utility by forming $\mathcal{N}_1$. Let β denote the net utility gain made by the member of $\mathcal{N}_1$ who attains the smallest gain if $\varepsilon = 0$ and $c = 0$. For $c = 0$ and ε sufficiently close to zero, members of $\mathcal{N}_1$ still attain a strict net gain in utility if they form $\mathcal{N}_1$. For an arbitrary $\lambda \in (0,1)$, let $\overline{\varepsilon} > 0$ be such that if $c = 0$, agents of $\mathcal{N}_1$ attain a net gain in utility of at least $\lambda\beta$ by forming $\mathcal{N}_1$ for any realization of preferences $\widetilde{x} \in N(x, \overline{\varepsilon})$. Then, for any $\varepsilon < \overline{\varepsilon}$, and preferences $\widetilde{x} \in N(x, \varepsilon)$, if $c < \lambda\beta$, agents in $\mathcal{N}_1$ achieve a strictly higher utility forming $\mathcal{N}_1$ than not forming $\mathcal{N}_1$. The uncovered set converges to the status quo as ε converges to zero, hence given a small enough ε, if $\mathcal{N}_1$ forms proposals pass outside the uncovered set.

2) As argued above, the correspondence of possible utilities given ε if a given set of voting blocs form converges as $\varepsilon \longrightarrow 0$ to the vector of utilities that the agents obtain if $\varepsilon = 0$ and this set of voting blocs form. For $\varepsilon = 0$, if c is low enough, given that A forms, B has an strict incentive to form, and given that B forms, A has an strict incentive to form. Since changes in utility converge to zero as ε converges to zero, it follows that for a small enough ε, if c is low enough, agents still have a strict incentive to form A and B. The same argument applies to A_1 and B_1 in case two.

3) Noting once again that changes in utility converge to zero as $\varepsilon \longrightarrow 0$, for a small enough ε, the agenda setter a can choose an agenda (ρ_1^t, ρ_2^t) on every odd issue (except T; if T is odd let $\rho^T = (0,0)$) and $(\rho_1^t, -\rho_2^t)$ on every even issue such that all agents with $x^i = (1,0)$ favor it. If c_a is low enough and agents in $\mathcal{N}_a$ form a unique voting bloc, this agenda passes and all the agents in set $\mathcal{N}_a$ strictly increase their utility. The best response of the agenda setter is to propose an agenda that leads to the formation of a voting bloc

and to a strict increase in utility for the agenda setter.

4) For a small enough ε, once agenda $\boldsymbol{\rho}$ is announced, bloc A strictly benefits from forming a unique voting bloc, so an outcome with no blocs is not an equilibrium with no coordination failures, just as in the proof of proposition 4; voting blocs must form and they must, with positive probability, affect the outcome, so ex ante a is strictly better off. It follows that the best agenda $\hat{\boldsymbol{\rho}}$ also generates a utility gain for a. The minimum possible utility gain is now a function of ε because as ε increases, a may benefit less from the proposals in agenda. Let $\beta(\varepsilon)$ denote this function. Suppose $\mathcal{N}_i$ forms a unique voting bloc. For an arbitrary $i \in B$, consider agent $j(i)$. For ε sufficiently small, if the agenda benefits the agenda setter by at least $\beta(\varepsilon)$ and it strictly benefits i, it must strictly hurt $j(i)$. Let $f(b, \varepsilon)$ be the minimum possible loss to $j(i)$ given outcomes that make i strictly better off and make a strictly better off by at least b. This function is decreasing in ε, and as $\varepsilon \longrightarrow 0$, it converges to a function of b that is bounded above zero by the same arguments as in the proof of 4. Thus, for ε sufficiently small, $f(x, \varepsilon) > 0$, and hence there exists $c_\varepsilon > 0$ such that given any $c < c_\varepsilon$, agent $j(i) \in B$ is strictly better off if B_2 forms. Find the lowest such threshold among all members of B. If the cost is lower than this threshold, every $j \in B$ strictly benefits if the bloc forms, so if the bloc does not form, it is a coordination failure. Hence in any equilibrium without coordination failures, at least two blocs form. ■

References

[1] Aldrich, J.H. *Why Parties? The Origin and Transformation of Political Parties in America.* Chicago, University of Chicago Press (1995).

[2] Ansolabehere, S., Leblanc, W., Snyder, J.M. When parties are not teams: party positions in single member district and proportional representation systems. *Econ. Theory [in this special issue].*

[3] Ashworth, S., Bueno de Mesquita, E. Informative party labels with institutional and electoral variation. *J. of Theor. Politics* 20(3): 251—273 (2008).

[4] Baron, D.P. A noncooperative theory of legislative coalitions. *American J. of Political Science* 33(4): 1048—1084 (1989).

[5] Baron, D.P., Diermeier, D., Fong, P. A dynamic theory of parliamentary democracy. *Econ. Theory [in this special issue].*

[6] Baron, D.P., Hirsch, A.V. Common agency lobbying over coalitions and policy. *Econ. Theory [in this special issue].*

[7] Besley, T., Coate, S. An economic model of representative democracy. *Q. J. of Econ.* 112(1): 85—114 (1997).

[8] Bianco, W.T., Sened, I. Uncovering evidence of conditional party government. *American Political Science* Rev. 99(3): 361—371 (2005).

[9] Caillaud, B., Tirole, J. Party governance and ideological bias. *European Econ. Rev.* 34: 779—789 (1999).

[10] Caillaud, B., Tirole, J. Parties as political intermediaries. *Q. J. of Econ.* 117(4): 1453—1489 (2002).

[11] Carraro, C. *The Endogenous Formation of Economic Coalitions*. Edward Elgar Publishing (2003).

[12] Carruba, C.J., Volden C. Coalitional politics and logrolling in legislative institutions. *American J. of Political Science* 44(2): 261—277 (2000).

[13] Cox, G.W. The uncovered set and the core. *American J. of Political Science* 31(2): 408—422 (1987).

[14] Cox, G.W., McCubbins, M.D. *Legislative Leviathan*. Berkeley, University of California Press (1993).

[15] Cox, G.W., McCubbins, M.D. *Legislative Leviathan: Party Government in the House*. Cambridge University Press (2007).

[16] Diermeier, D., Merlo, A. Government turnover in parliamentary democracies. *J. of Econ. Theory 94(1): 46—79 (2000).*

[17] Diermeier, D., Vlaicu, R. Parties, coalitions, and the internal organization of legislatures. SSRN w.p. 1150963 (2008).

[18] Eguia, J.X. Voting blocs, party discipline and party formation. SSRN w.p. 1514506 (2010).

[19] Eguia, J.X. Endogenous parties in an assembly. *American J. of Political Science* 55(1): 16—26 (2011).

[20] Fox, J. Legislative cooperation among impatient legislators. *J. of Theor. Politics* 18(1): 68—97 (2006).

[21] Hart, S., Kurz, M. Endogenous formation of coalitions. *Econometrica* 51(1): 1047—1064 (1983).

[22] Jackson, M., Moselle, B. Coalition and party formation in a legislative voting game. *J. of Econ. Theory* 103(1): 49—87 (2002).

[23] Koford, K.J. Centralized vote-trading. *Public Choice* 39(2): 245—268 (1982).

[24] Krehbiel, K. *Pivotal Politics: A Theory of U.S. Lawmaking.* University of Chicago Press (1998).

[25] Levy, G. A model of political parties. *J. of Economic Theory,* 115(2): 250—277 (2004).

[26] McKelvey, R.D. Covering dominance and institution free properties of social choice. *American J. of Political Science* 30(2): 283—314 (1986).

[27] Morelli, M. Party formation and policy outcomes under different electoral systems. *Rev. of Econ. Stud.* 71(3): 829—853 (2004).

[28] Nash, J. Equilibrium points in n-person games. *Proc. of the Natl. Acad. of Sciences* 36: 48—49 (1950).

[29] Osborne, M., Tourky, R. Party formation in single issue politics. *J. of the European Econ. Assoc.* 6(5): 974—1005 (2008).

[30] Plott, C.R. A notion of equilibrium and its possibility under majority rule. *American Econ. Rev.* 57(4): 331—347 (1967).

[31] Ray, D. *A Game-Theoretic Perspective on Coalition Formation.* Oxford University Press (2007).

[32] Romer, T., Rosenthal, H. Political resource allocation, controlled agenda, and the status quo. *Public Choice* 33(4): 27—44 (1978).

[33] Schofield, N., Sened, I. *Multiparty Democracy.* Cambridge University Press (2006).

[34] Snyder, J., Ting, M. An informational rationale for political parties. *American J. of Political Science* 46: 90—110 (2002).

[35] Stratmann, T. The effects of logrolling on congressional voting. *American Econ. Rev.* 82(5): 1162—1176 (1992).

[36] Von Neumann, J., Morgenstern, O. *Theory of Games and Economic Behavior.* Princeton, Princeton University Press (1944).