A Spatial Theory of Party Formation.*

Jon X Eguia†

New York University

February 17, 2009

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.

Keywords: Party formation, coalition formation, policy making.

JEL classification: D71, D72.

*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.

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

Members of an assembly that chooses policies by majority voting have strategic incentives to coalesce and coordinate their votes, forming political parties.

I present a non cooperative theory of endogenous party formation in which any set of heterogeneous members of the assembly can coalesce, by unanimous agreement, into a voting bloc. Agents who coalesce exercise party discipline, casting their votes together in the assembly. I interpret these blocs as political parties.

I find that if there exists an agent who enjoys an advantage and can coordinate and organize a political party at a lower cost, then a single political party forms around this agent. If all agents face the same cost of forming political parties, at least two political parties form in every equilibrium, and an equilibrium with exactly two parties exists. In this equilibrium, two parties form at opposite sides of each other in the space of policy preferences.

I study a legislative assembly that makes a sequence of policy decisions, each of them multidimensional. Members of the assembly have Euclidean preferences around their ideal point in the policy space. I assume that at a cost, agents can coalesce into endogenous voting blocs to coordinate their votes. A Condorcet winner, if it exists, is a policy that defeats any other in pairwise comparisons. A status quo is the policy that is implemented if no other policy gathers a majority of votes in the assembly. In standard theories of policy making, if the status quo is a Condorcet winner, 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 legislators whose preferences are similar in one dimension, but diverge in a second dimension, have an incentive to coalesce into a voting bloc, committing to vote together on every issue 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 policy outcome moves away from the status quo, to the advantage of the members of the bloc.

My theoretical contribution to the non cooperative coalition formation literature is to study the formation of coalitions that generate both positive and negative ex-

ternalities to other agents. Traditional models of coalition formation assume that agents only care about the coalition they belong to. This is a very restrictive assumption, that rules out any externality across coalitions. The *partition function* approach pioneered by Lucas and Thrall (1963) assumes that the utility of an agent depends on the whole coalition structure in the society. Carraro (2003), Ray (2007) and Humphreys (2008) survey the recent progress of this literature. Bloch (2003), Yi (2003) and Bloch and Gomes (2006) study coalitions that generate either positive externalities or negative externalities, but not both. Hyndman and Ray (2007) allow for both types of externalities in a model with three agents. In this paper I consider an application with any finite number of agents in which coalitions generate both types of externalities: The endogenous formation of political parties in an assembly. Political parties benefit other agents who have similar preferences, and hurt agents with opposed preferences.

Substantively, this paper is complementary to other accounts 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 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, which Duverger (1959) terms "parties of parliamentary origin." 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. Both Baron (1989) and Jackson and Moselle (2002) seek to explain the formation of a ruling coalition that distributes pork. I seek instead to explain the incentives to form parties to affect the policy outcome in a purely ideological space of preferences. These parties need not be of majority size. In case studies of US political parties, Cox and McCubbins (1993) find that legislators in the majority party use the party to control the agenda, and Aldrich (1995) explains that US parties serve to coordinate a durable majority to reach a stable policy outcome avoiding majority cycles. My theory provides a different explanation of party formation that is robust to the consideration of both an exogenous and an endogenous agenda, and to distributions of individual preferences such that the social preference is acyclic as well as distributions that allow for cycles in the social preference.

In a related paper on party formation, Eguia (2006), I consider a binary policy space and I show that legislators with probabilistic preferences have incentives to coalesce into voting blocs. While in both papers the incentive to form voting blocs is to achieve a more desirable policy outcome, in the current manuscript I consider a multidimensional policy space, which allows for a richer preference relation than a mere binary space, and I also study the effects of endogenizing the agenda.

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 issues can benefit every member if the identity of the minority and majority within the party changes across issues.

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." This is the view I embrace. In the words of Craig and Volden: "The groundwork is now laid for a future analysis of parties." 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, of transnational parties. Diermeier and Merlo (2000) and Schofield and Sened (2006) among others 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. For instance, in the European Parliament, where there is no government to be formed, national parties have merged into pan-European parties. Hix, Noury and Roland (2007) report that

> "the cohesion of parties in the European Parliament has increased as the powers of the Parliament have increased. The authors suggest that the main reason for these developments is that like minded MEP's have incentives to form stable transnational party organizations to compete over European Union policies."

I provide a theoretical account of these incentives. After an illustrating example,

I present the theory, and then I first show my results on party formation 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 agents who make two policy decisions, each of which is two dimensional. That is, on each of two issues, the assembly must choose an outcome in a policy space with two dimensions. The status quo on each issue is $(0, 0)$. The ideal policy of agent 0 is at the status quo for both issues. 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 issue 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 issue separately. The agenda on each issue puts to a vote the status quo versus a policy proposal randomly chosen from the Pareto set of policies $[-1, 1]^2$. The policy proposal passes if five or more agents vote for it.

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 agent 4 is worse off. Ex ante, it is equally

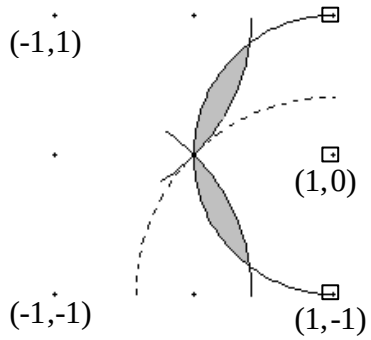

Figure 1: A voting bloc forms and changes the policy outcome.

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 indifferent 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 a legislative assembly $\mathcal{N}$with N (odd) agents who make a decision on each of T different policy issues. Each policy issue is two dimensional and the status quo policy on each issue is $(0,0)$. On each issue, a policy proposal is put to a vote and if it gathers a simple majority of votes, it becomes the policy outcome on that issue, otherwise the status quo remains in place. The assembly then moves on to decide on the next issue. Legislators have circular preferences around their ideal policy, and their utility is additive across issues, without discount. For each legislator $i \in \mathcal{N}$, let $p_i \in \mathbb{R}^2$ be the ideal policy of the legislator i, constant for every issue, let $p^t \in \mathbb{R}^2$ be the policy outcome on issue t, and let $p = (p^1, ..., p^t, ..., p^T)$ be the vector of policy outcomes. Let $\|\cdot\|$ be the Euclidean norm, then $u_i(p) = \sum_{t=1}^{T} - \|p^t - p_i\|$.

The ideal policies of the legislators are distributed on a grid around the origin, and the distribution is symmetric with respect to both axes. Formally, for some finite K, let the size of the grid be $2K \times 2K$ and for any $a, b \in \{-K, -K+1, ..., K-1, K\}$, let $N_{a,b}$ denote the number of agents with ideal policy (a,b). Then the symmetry assumption is that $N_{a,b} = N_{-a,b} = N_{a,-b} = N_{-a,-b}$. Further, I assume that $N_{x,y} \geq 1$ for any x, y such that $\max\{|x|, |y|\} \leq 1$, 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.

The size of the grid is arbitrary. If there is at least one legislator with an ideal policy at any given point on the grid, $K = 5$ generates an assembly larger than the US Senate, and $K = 10$ larger than the US House of Representatives. A finer grid approximates arbitrarily close any 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 legislators vote their true preference in the assembly, every proposal fails.

I assume that, at a cost, legislators can make binding commitments to coordinate their votes on all issues. The timing 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, which is a $2 \times T$ matrix specifying a two dimensional policy for each issue. If the agenda is exogenous, this step is skipped, and agents know that Nature chooses the agenda at a later stage. The proposed agenda is revealed and becomes public knowledge.

Second, every agent can issue a proposal or invitation to any subset of other agents to form a voting bloc that includes the proposer. These proposals all become common knowledge as well.

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 legislator who receives a proposal to form a given voting bloc accepts it, the bloc forms. Organizing this bloc is costly, and every member, including the legislator who first made the proposal, bears a cost $c > 0$. In some instances I will consider a lower level of cost $c_a < c$ for a bloc proposed by the agenda setter. Agents join a bloc strategically, joining only if it makes them better off.

Fourth (only if the agenda is exogenous), Nature chooses it by independently drawing a point for each issue from a distribution that is uniform in $[-1,1]^2$. The chosen agenda becomes public and common knowledge.

Fifth, legislators who are members of a bloc meet on a caucus and they vote on each policy issue, choosing between the policy proposal detailed in the agenda or the status quo. Every voting bloc coordinates by simple majority: If a simple majority of its members votes in favor of the proposal on a given issue in the caucus, the bloc as a whole 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, issue by issue, 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.

I assume that legislators 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. If the agenda is endogenous, the strategy of the potential agenda 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 issue T is weakly dominant on the last subgame. On issue $t' < T$, if on every issue $t > t'$ agents vote sincerely, then by backward induction it is weakly dominant to vote sincerely on issue 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 protocol to form a voting bloc is similar to Hart and Kurz's (1983) coalition game Γ, first introduced by von Neumann and Morgenstern (1944). 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.

The first result is a partial equilibrium result, solving the game in which only one legislator has the ability to send invitations to form a voting bloc. Let agent l have ideal policy $p_l = (x_l, y_l)$ such that $x_l \neq 0$ and $y_l = 0$.

Proposition 1 *Let the agenda be exogenous and assume that only legislator l can issue invitations to form a voting bloc. An equilibrium exists. If the cost c is low enough, in every equilibrium a voting bloc forms and the policy proposal on every issue passes with positive probability.*

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)) or to eradicate cycles and 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 that exercises party discipline, purely for ideological gain, even if they have no control over the agenda, and even in the absence of majority cycles or stability. In Proposition 1 I show that a set of agents who coordinate their votes forming a voting bloc succeeds in defeating a Condorcet winning policy, and they are able to move the policy outcome in a way that is favorable to the bloc.

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. The discretion of an agenda-setter with positive agenda control is proportional to the distance between the Condorcet winner (again, the median in one dimension) and the status quo in the seminal agenda-setter theory by Romer and Rosenthal (1978). 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 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 and toward their preference.

Members of a bloc exploit their common preference in one dimension, and they obviate their conflicting preferences on a second dimension, agreeing to vote for policies that bring a desired change in the dimension they agree upon. In this manner,

they defeat the status quo policy with positive probability. Note that members of a bloc benefit in expectation. Ex post, there is a net aggregate gain for the bloc, but some members may be worse off. If the number of issues put to a vote 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 issue put to a vote.

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 assuming that forming a vote is costly, that the bloc cannot control the agenda and that there are no majority cycles to exploit, a disciplined 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 legislator can invite others to form a voting bloc. In expectation, some non members become worse off. If they can form their own voting blocs, they too have incentives to coalesce. If legislators receive more than one invitation to join a bloc, coordination issues arise. For instance, if legislators i and i' both issue invitations to 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 forms 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 in view

of what other agents do. The expectation is with respect to the realization of the agenda if it is exogenous, and 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 *Let the agenda be exogenous and let any legislator be able to propose forming a voting bloc. There is no equilibrium without coordination failures with a unique voting bloc. If c is low enough, there exists 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 parties is constructive. In the equilibrium I describe, agents separate into parties 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 a legislature which some legislators 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. Assume that any legislator can propose the formation of a party at a cost c to each of its members, but the agenda setter can organize a party at a cost c_a to its members, with $c_a \leq c$. A difference in cost allows the possibility that the agenda setter enjoys an advantage in her ability to organize and coordinate other legislators. Perhaps the agenda setter has control over the executive branch of government, or over bureaucratic appointments, so she has carrots and rewards to offer to legislators that join the voting bloc, while these tools to foster discipline and coordination are not available to the opposition. Other legislators may be at a disadvantage, and face greater difficulties to secure commitments: Maybe they need to make deposits or bonds that would be lost by defectors, or they need to create a tight social network that could impose social sanctions to defectors before they can sign credibly binding commitments. Irrespective of the causes, I consider the possibility that some legislators face greater difficulties to make binding commitments than others.

If the cost of coordination is lower for the agenda setter, she is able to exploit her privileged position for political gain. She forms a voting bloc and proposes a sequence of policy proposals that benefit every member of her bloc. Agents with the opposite preferences can render this bloc ineffective by forming their own bloc. The agenda setter can prevent this outcome by crafting an agenda that is beneficial to the members of her bloc, but 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 and agent r with ideal policy $p_r \neq (0,0)$ is the agenda setter. Given any cost of coordination c, if the cost for blocs proposed by the agenda setter $c_a > 0$ is low enough, at least one voting bloc forms with positive probability and the policy outcome moves away from the status quo, benefiting the agenda setter.*

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 policy proposals on each issue 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 policy 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: (x, y) along the indifference curve of the agents with ideal policy $(-1, 1)$ in all the odd issues and $(x, -y)$, which lies along the indifference curve of the $(-1, -1)$ agents, in all even issues, with $x, y > 0$, and choosing (x, y) just close enough to $(0, 0)$ so that no other legislators have an incentive to propose a second voting bloc. That is, the agenda proposes the most favorable pair of points in the shaded areas in figure 1 that are symmetric to the x axis and that do not lead the opposition to form a second bloc. In every issue, the agenda setter proposes to move the policy to the right. In half the issues 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 who do not benefit from these proposals by 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.

A discussion of an example by Schwartz (1977) best illustrates the basic insight behind Proposition 3. Paraphrasing Schwartz: "An organized minority has frustrated the wishes of a disorganized majority. Had [other agents] also traded votes, they would have blocked the effect of the first trade. If cooperation costs are such as to permit [some agents] to trade votes while preventing [other agents] from trading them, then the will of the majority is frustrated."

The result in Proposition 3 goes beyond vote trading. Vote trading typically consists on identifying a pair of votes such that two or more legislators agree to trade support in one vote for support on the other vote so that both votes pass, and all legislators involved in the trade become better off. The voting blocs in my theory play a more ambitious role: Legislators commit to coordinate all their future votes, always casting them together. They may do so uncertain about the agenda they vote on, as in propositions 1 and 2, or they may devise and coordinate around an agenda that they all vote for, as in proposition 3. Legislators coalesce as a voting bloc and pass a series of policies that collectively make them all better off, while keeping a potential opposition disorganized, by choosing an agenda that would not prompt the opposition to coordinate. I find it easier to interpret the actions of legislators who endogenously formulate an agenda and who commit to all support this agenda exercising party discipline as the workings of an emerging political party than as a mere exercise in vote trading. In this light, Proposition 3 shows that at least one political party forms endogenously around the agenda setter for ideological gain, aided by built-in advantage of being able to coordinate at a lower cost than the opposition.

If no legislator has an advantage in organizing voting blocs, no single voting bloc can prosper and move the policy outcome. 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 at every stage, 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.[1] 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. The most straightforward explanation of this apparently self-hurting behavior is that the mere act of fighting, of forcing a vote and being on record on the losing side, provides a direct benefit irrespective of policy outcomes, perhaps because the electorate rewards the mere act of fighting for ideas, regardless of success in terms of implemented policy.

I consider an alternative explanation for aggressive agendas and party discipline that does not rely on the expressive value of voting for a losing proposition. 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 probability λ, coordination fails and commitments are not binding. Agents assume the same cost in joining a voting bloc. However, now this cost is not a sure investment but instead, a risky one: With probability λ, despite the effort at coordinating, the voting bloc fails, commitments are not binding, there is no party discipline 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,

[1] These are roll calls 44, 51, 75, 167, 241, 243, 252, 341, 345, 362, 411, 437 and 438. In ten of these roll calls, between 47 and 56 out of a hundred senators voted in favor along party lines, but in some cases, 60 votes were necessary for passage.

parties have an incentive to propose an aggressive agenda in the hope that their party succeeds in coordinating, and the opposition does not.

Proposition 4 *Assume* $T \geq 2$, *the agenda is endogenous, the potential agenda setters are the agents with any ideal policy* $(x, 0)$ *such that* $x \neq 0$, *and there is uncertainty* $\lambda \in (0, 1/2)$ *about the enforcement of commitments. If the cost* c *of coordinating a bloc is low enough, in any equilibrium without coordination failures at least two voting blocs form.*

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 policy proposal succeed(s) in enforcing party discipline while the other bloc(s) fail(s). In this case, the agenda passes in the assembly and the policy proposal moves 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 $(x_a, 0)$ and $(x_b, 0)$ such that $x_b = -x_a$, 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 policy is a Condorcet winner to show that voting blocs form and change the policy 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 policy and there is a set of policies that are preferred by a majority of legislators over the status quo. However, each of these policies 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). A policy x *covers* y if x beats y and any policy z that beats x also beats y according to majority preferences. The policies 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 of party formation shows that voting blocs form and move the policy outcome outside the uncovered set.

Let us disturb the preferences. Given the original configuration of ideal policies p_i for each $i \in \mathcal{N}$, which satisfy radial symmetry, for each $i \in \mathcal{N}$, let $\widetilde{p}_i$ be the new ideal policy of the agent, such that $\widetilde{p}_i \in N(\widetilde{p}_i, \varepsilon)$, where $N(p, \varepsilon)$ is the neighborhood of size ε around p. Let $\widetilde{p}$ be the $N \times 2$ matrix that denotes the ideal policy of every agent, and, along with the assumption of Euclidean preferences, determines the preferences of every agent. This new preference profile is more general, since it relaxes radial symmetry. If the disturbed preferences are close enough to the original preferences on a grid, propositions 1-4 hold, subject to the appropriate restatement.

Proposition 5 *Let the agents hold preferences given by* $\widetilde{p}$.

1) Assume the agenda is exogenous and only legislator l *with* $p_l = (x_l, 0)$, $x_l \neq 0$ *can issue invitations to form a voting bloc. There exists* $\overline{\varepsilon} > 0$ *such that for any* $\varepsilon \leq \overline{\varepsilon}$,

if c is low enough, in every equilibrium a voting bloc forms. With positive probability, the policy proposal in every issue defeats the status quo and the policy outcome moves outside the uncovered set.

2) Assume that the agenda is exogenous and any legislator is able to propose forming a voting bloc. There exists $\overline{\varepsilon} > 0$ such that for any $\varepsilon \leq \overline{\varepsilon}$, if c is low enough, 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 and i with ideal policy $\widetilde{p}_i \neq (0,0)$ is the agenda setter. Given any $c > 0$, there exists $\overline{\varepsilon} > 0$ such that for any $\varepsilon \leq \overline{\varepsilon}$, if $c_a > 0$ is low enough, at least one voting bloc forms with positive probability and the policy outcome moves away from the uncovered set, benefiting the agenda setter.

4) Assume $T \geq 2$, the agenda is endogenous, the potential agenda setters have ideal policies $p = (x, 0)$ such that $x \neq 0$, and there is uncertainty $\lambda \in (0, 1/2)$ about the enforcement of commitments. There exists $\overline{\varepsilon} > 0$ such that for any $\varepsilon \leq \overline{\varepsilon}$, if c is low enough, 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 create political parties 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, or in the control of the agenda, a unique party that benefits this agent forms. Otherwise, at least two parties form.

Discussion

I have presented a theory of endogenous party formation in which members of a legislative assembly who have ideological preferences on a multidimensional policy space coordinate to form voting blocs. These voting blocs function as strong political parties that exercise party discipline so that all party members cast their votes together in the assembly.

I have shown that if one agent has an exogenous technological advantage that allows her to coordinate others at a lower cost, in equilibrium one party forms around this agent, and the policy outcome moves away from the status quo policy, 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 parties that cannot enforce party discipline. In this case, the outcome of the internal meeting of the party is only a voting prescription that the members of the party can follow or ignore in the assembly, without punishments for ignoring it. If there is no utility loss of any kind from ignoring the prescription, following it against sincere preferences is a weakly dominated strategy. Suppose instead that a party member who deviates from the voting prescription of her party incurs an arbitrarily small cost $\varepsilon > 0$. With undominated strategies, party coordination can then affect the policy 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 party member has an individual incentive to deviate from the party's voting prescription. Parties 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 that account for the impossibility of coordination in close votes, party members have an incentive to vote according to their party's voting prescription in all other votes, and every agent votes sincerely in the assembly in cases when coordination would result in an outcome decided by one vote. At the internal meetings, all agents have an incentive to vote sincerely. Therefore, without commitment, parties become coordination devices for policy decisions where no agent is individually pivotal, and they still affect some policy outcomes and benefit their members in expectation, so agents have (almost) the same incentives to form parties as in the case with binding commitment.

My theory of party formation provides an explanation for the endogenous emergence of political parties inside an assembly, without the intervention of outside actors such as an electorate. Unlike other models of party formation inside an assembly, such as Baron (1989) and Jackson and Moselle (2002), my theory does not rely on bargaining over distributive policies. The prediction of bargaining models in which agents vote over the allocation of a fixed amount of money is for a unique coalition to form, of minimal winning size. My theory, based on ideological preferences over a multidimensional policy space, predicts instead that political parties of size less than minimal winning form, and two or more parties coexist. The intuition for this different prediction is simple: In distributive policies, agents who are excluded from any share of the pie that is to be divided always oppose the division, so a coalition must be big enough to win by itself, and the losing minority has no incentives to form its own coalition. On the other hand, if policy preferences are ideological, agents who do not belong to a party may still benefit and vote for the policy proposals that the party favors. Minority parties can attain sufficient external support to move the policy outcome to their gain. Other minority parties with the opposite preferences have incentives to

form as well to prevent undesired policy proposals from passing.

The formation of a party benefits party members, but the net effect of the formation of parties is a loss of social welfare. The findings in this paper have relevance for the literature on constitutional design and the selection of voting rules, in particular, for the desirability of majority rule. I have considered an assembly in which the status quo policy is the socially optimal policy and it is majority preferred to any other policy, so that if agents cast their votes in the assembly according to their sincere preference, majority rule ensures a socially optimal policy outcome. I show that if a unique voting bloc forms in equilibrium, the vote tally in the assembly does not reflect true preferences and the status quo is defeated in favor of policies preferred by the bloc, generating a loss in utilitarian social welfare. If the assembly required instead a supermajority of votes in order to implement any change in policy, the socially optimal status quo would be more easily preserved. The traditional normative argument in favor of supermajority rules is that they protect minorities from exploitation by a tyranny of the majority. My results suggest that in some circumstances, a supermajority rule is necessary to protect the majority of the assembly from exploitation by an organized minority that strategically coordinates its votes.

The strategic collusion of agents who coordinate their actions has wider implications for mechanism design. A benevolent social planner that lays out a mechanism for the aggregation of preferences into policy outcomes must anticipate that, given a mechanism, agents form equilibrium coalitions to coordinate their actions. Mechanisms, or specifically voting rules, that are optimal given individual incentives may not be desirable given the formation of coalitions and parties in equilibrium. The design of optimal voting rules that take into account the incentives of agents to form coalitions and parties to coordinate their actions is an exciting research agenda.

Appendix

Proposition 1

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

For any x, y, let $i_{x,y}$ denote an arbitrary agent with ideal policy $p_i = (x, y)$. Without loss of generality, let $x_l > 0$. Let $A_1 = \{i_{x,y} : x > 0 \text{ and } x - 1 \leq |y| \leq x\}$. Consider the following strategy: Agent l proposes the formation of $A = l \bigcup A_1$. On each issue, the bloc A favors the policy proposal p_t if and only if $i_{1,0}$ favors it. If $i_{1,0}$ favors it, l favors it and either all $i_{x,y} \in A$ with $y > 0$ or all $i_{x,y} \in A$ with $y < 0$ favor it as well, constituting a majority of the bloc in favor. If $i_{1,0}$ prefers the status quo, either all $i_{x,y} \in A$ with $y > 0$ or all $i_{x,y} \in A$ with $y < 0$ prefer the status quo as well. Given that A never votes as a bloc against the preference of $i_{1,0}$, no policy gathers a majority in the assembly if $i_{1,0}$ opposes it. Given a policy p_t, if $i_{1,0}$ and $i_{-1,1}$ and $i_{x,y}$ favor it for all x, y such that $x \geq 2$ and $y = -x + 2$, then p_t passes in the assembly. Similarly, p_t passes if $i_{1,0}$ and $i_{-1,-1}$ and $i_{x,y}$ favor it for all x, y such that $x \geq 2$ and $y = x - 2$. Since the slopes of the indifferent curves of $i_{1,0}$ and $i_{x,y}$ such that $x \geq 2$ and $y = -x + 2$ 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 policies 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 policies that pass if A forms is divided into two areas, symmetric with respect to the horizontal axis. I illustrate the two sets of policies that pass, shaded, in figure 1 for an assembly with $K = 1$ and $N_{a,b} = 1$ for $a, b = \{0, 1\}$, so $N = 9$, and with $p_l = (1, 0)$. In this case, policies pass if and only if $i_{1,0}$ and either $i_{-1,1}$ or $i_{-1,-1}$ favor them.

I want to show that for any pair of policies (a, b) and $(a, -b)$ such that $i_{1,0}$ prefers this pair of policies to pass and be implemented, better than the pair of policies $(0, 0)$ and $(0, 0)$, every member of A prefers (a, b) and $(a, -b)$ as well. Since any such pair that makes $i_{1,0}$ strictly better off is better for any $j \in A$ than another pair with the

same b, but with the first coordinate moved left to the point where $i_{1,0}$ is indifferent, it suffices to show that every $j \in A$ weakly prefers a pair of policies that make $i_{1,0}$ indifferent over the status quo implemented twice. A policy (a, b) makes $i_{1,0}$ indifferent if and only if

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

Formally, I want to show that for any (x, y) such that $x > 0$ and $-x \leq y \leq x$, and any (a, b) such that $a = 1 - \sqrt{1-b^2}$,

$$[(x-a)^2 + (y-b)^2]^{1/2} + [(x-a)^2 + (y+b)^2]^{1/2} \leq 2[x^2 + y^2]^{1/2}.$$

Algebraic manipulations (omitted, but available from the author) of this expression yield

$$2(x^2+y^2) + 2(x^2+y^2)(x-1)\sqrt{1-b^2} \leq b^2y^2 + 2x^3 + 2xy^2$$

Since $0 \leq b \leq 1$ and $x \geq 1$, it suffices to check

$$2(x^2+y^2)x \leq b^2y^2 + 2(x^2+y^2)x$$

which is always true for any y. Furthermore, if $b \neq 0$ the left hand side is strictly less than $2(x^2+y^2)x$ and the last expression holds with strict inequality.

Given any point (a, b) that passes if A forms, $(a, -b)$ passes as well. Given the uniform density function of the exogenous agenda, the density function of the policy proposals on issue t is the same at policies (a, b) and $(a, -b)$. It follows that every $i \in A$ strictly benefits from the coordination of votes inside A. If c is low enough, the benefit outweighs c and it is a dominant strategy for every agent who receives an invitation to join the bloc. Then, agent l strictly gains ex ante from issuing this invitation.

Consider the subgame after l makes any other 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. For c low enough, at least one subgame yields a payoff gain strictly higher than c. The maximum payoff gain then is also strictly greater than c. An outcome does not generates a strictly positive payoff gain unless the policy proposal defeats the status quo with positive probability, which occurs only if a voting bloc forms. It follows that the best response 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. Ex ante all members of the bloc become better off. ■

Proposition 2

Proof. I first rule out equilibria without coordination failures and with no blocs. As shown in the proof of proposition 1, $\exists A \subset \mathcal{N}$ such that if $c > 0$ is low enough, every $i \in A$ is strictly better off if A forms a unique voting bloc. Hence 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 i strictly benefits from the formation of A. Pair all agents with $p_i \neq (0,0)$ as follows: For any i with $p_i = (a,b)$, let $j(i) : \mathcal{N} \longrightarrow \mathcal{N}$ be a one-to-one mapping such that $p_{j(i)} = (-a,-b)$. For any i and $j(i)$, if i strictly benefits from a change in policy from $(0,0)$ to p, then $j(i)$ is hurt at least as much i benefits from this change. Let $B = \{j(i) : i \in A\}$. If A and B each forms a bloc and no other agent forms another bloc, they cancel each other out and no policy proposal passes. Hence if A has an incentive to form a voting bloc, then the set of agents in B have an incentive to form a voting bloc as well, 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 $A_1 = \{i : p_i = (x,y)$ with $x > 0\}$, let $B_1 = \{j : p_j = (x,y)$ with $x < 0\}$, and let $C = \{i : p_i = (0,0)\}$. Recall $i_{x,y}$ denotes an agent with ideal policy (x,y). I consider two separate cases.

One– Either $N_{0,0} > 1$, or $\sum_{b=1}^{K} N_{0,b}$ is even. Let ρ be an ordering of every agent with ideal policy $(0, y)$, with $y > 0$, and let $\rho(i) \in \{1, 2, ...\}$ be the position of i in order ρ. Let ρ be such that for any i, j with $p_i = (0, y_i)$ and $p_j = (0, y_j)$, then $y_j > y_i$ implies $\rho(j) > \rho(i)$. Let $A_2 = \{i : \rho(i) \text{ is odd}\}$ and let $B_2 = \{i : \rho(i) \text{ is even}\}$. Let $A_3 = \{j(i) : i \in A_2\}$ and let $B_3 = \{j(i) : i \in B_2\}$. Let $A = A_1 \cup A_2 \cup A_3$ and let $B = B_1 \cup B_2 \cup B_3$. Note that $\{A, B, C\}$ partitions the assembly and $|A| - |B| = 0$ if $\sum_{b=1}^{K} N_{0,b}$ is even, or $|A| - |B| = 2$ otherwise; in either case $|A| - |B| < N_{0,0}$ and $|B| \leq |A| < N/2$.

Consider the following strategy. At the proposal stage, $i', i'' \in A$ propose the formation of A and $j', j'' \in B$ propose the formation of B. No other agent other than these four proposes forming a voting bloc. At the acceptance stage, given these four proposals, all $i \in A$ including i'' accept the invitation by i', and all $j \in B$ including j'' accept the invitation by j'. If i' and/or j' deviate at the proposal stage, agents accept the invitations by i'' and j'' instead. All agents in $A \cup B$ ignore any other individual deviation at the proposal stage. Agents in C ignore any invitation at the proposal stage, unless it is weakly dominant for them to accept it, in which case they accept it. Following deviations by more than one agent at the individual stage, agents play an arbitrary undominated equilibrium of the continuation game.

With these acceptance strategies, no agent has an incentive to individually deviate at the proposal stage, since the outcome does not change following the deviation. At the acceptance stage, assuming that the acceptance strategies for the equilibrium proposals are indeed a best response, after a deviation at the proposal stage that proposes the formation of a voting bloc D with at least two members of $A \cup B$, it is a mutual best response for every invited agent to reject this invitation to join D, because given that the other agent(s) reject(s) the invitation to form D, D is not going to form. Accepting an invitation to form a bloc D that only includes one member $i \in A \cup B$ is not a best response for i because all agents outside $A \cup B$ have the same ideal policy, so only i has her vote reversed by party discipline in D. Given that the

distribution of ideal policies inside A and B are symmetric with respect to the x axis, the set of policies that pass if either A or B forms is symmetric with respect to the x 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 policies that pass if only A forms is symmetric to the x axis and completely to the right of the y 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.

Two– $N_{0,0} = 1$ and $\sum_{b=1}^{K} N_{0,b}$ is odd. At the proposal stage, $i', i'' \in A_1$ propose the formation of A_1, $j', j'' \in B_1$ propose the formation of B_1, and $i_{0,0}$ proposes the formation of $C_1 = \mathcal{N} \backslash \{A \cup B\} = \{i : p_i = (x, y) \text{ with } x = 0\}$. No other agent other than these five proposes forming a voting bloc. At the acceptance stage, given these four proposals, all $i \in A_1$ including i''accept the invitation by i', and all $j \in B_1$ including j'' accept the invitation by j' and all members of C_1 reject the invitation to form C_1. If i' and/or j' deviate at the proposal 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 proposal stage and continue to accept the invitations by i' and j'. Agents in C_1 ignore deviations at the proposal stage that do not involve members of C_1 and continue to reject the invitation to form C_1. These agents also ignore proposals that invite members of C_1, as long as ignoring them is not undominated. If they receive an invitation that weakly dominates rejection of all invitations, then all $i \in C_1$ except $i_{0,1}$ accept the invitation to join C_1, and $i_{0,1}$ accepts the new invitation.

With these acceptance strategies, no agent has an incentive to individually deviate at the proposal stage, since the outcome does not change following the deviation. It remains to be shown that these acceptance strategies are best responses. By the same argument as in case one, if A_1 forms and c is low enough, every $i \in B_1$ is better off if B_1 forms, hence accepting the invitation to form B_1 is a best response, and an analogous argument applies to A_1. 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. As before, 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 $A_1 \cup B_1$ and this invitation is such that it makes ignoring all invitations weakly dominated, given that only 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 the members of C_1, rejecting both the invitation to form C_1 and the invitation to form D is weakly dominated, but accepting either, if c is low enough, is 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, $i_{0,1}$ accepting D and all others accepting C_1 is a mutual undominated best response. Hence the described strategy profile is an undominated equilibrium. ■

Proposition 3

Proof. Let r have an ideal policy $p_r = (x_r, y_r)$ such that $x_r > 0$ and $|y_r| \leq x_r$ (an analogous proof applies to any of the other seven sections where p_r may lie). Suppose r proposes the following agenda a_r: (x_a, y_a) in every odd issue except T (if T is odd), and $(x_a, -y_a)$ in every even issue, with $x_a > 0$ and (x_a, y_a) lying on the indifference curve of the agents with ideal policy at $(-1, 1)$ that runs through the status quo (figure 1 depicts this indifference curve), and at a point such that $i_{1,0}$ prefers (x_a, y_a) better than $(0,0)$. Suppose r proposes the formation of the bloc $A = \{i_{x,y} : x > 0, |y| \leq x\}$. If A forms, the agenda passes and, as shown in the calculations of proposition 1, every $i \in A$ strictly benefits. Given c, if (x_a, y_a) is close enough to the status quo, 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 r. If the cost c_a of joining blocs proposed by the agenda setter is low enough and the agenda setter proposes A, every $i \in A$ strictly benefits if A forms and after eliminating any strategy profile in which other blocs form, it is weakly dominated not to join A, so every $i \in A$ joins it, the policy outcome changes in every issue and every $i \in A$ becomes strictly better off.

Consider all the subgames that follow after r announces any other agenda. All these subgames are finite, hence they have an equilibrium. Take any set of equilibrium strategy profiles, one per 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 $p_r = (x_r, 0)$ with $x_r > 0$ be the chosen agenda setter. Let $A = \{i_{x,y} : x > 0\}$.

Suppose r proposes $a_r^1 = (0.1, 0.25)$ in every odd t except $t = T$ (if T is odd), $(0.1, -0.25)$ in every even t, and $(0, 0)$ at $t = T$ if T is odd. Every $i \in A$ strictly prefers this agenda to pass, better than keeping the status quo in every issue. If a_r^1 is the chosen agenda and A forms a unique voting bloc, p^t passes in every t. It follows that, for c low enough, if a_r^1 is the chosen agenda and no voting blocs form, A is in a coordination failure in the subgame after a_r^1 becomes the agenda. Let $V = \{V_1, ..., V_J\}$ be the set of voting blocs that form after a_r^1 becomes the agenda. As argued, in any equilibrium with no coordination failures, V is non empty. Take an arbitrary $V_j \in V$

and let $V = V_j \bigcup V_{-j}$. In order for the members of V_j to be willing to form this bloc, it must be that, with some probability, the enforcement of commitments by V_j alters the policy outcome given the formation and probabilistic enforcement of commitments by the blocs in V_{-j}. Therefore, given that the blocs in V_{-j} form, either the policy proposals in a_r^1 pass with positive probability if V_j does not form, or they pass with positive probability if V_j forms. Since V_j may fail to enforce commitments even if it forms, it follows in either case that the policy proposals in a_r^1 pass with positive probability if V forms. Thus, in any equilibrium without coordination failures of the subgame after a_r^1 becomes the agenda, the agenda setter r is in expectation strictly better off than in the benchmark with the status quo outcome and no voting blocs. If there are multiple equilibria, let β be a lower bound on the utility that the agenda setter gains by proposing a_r^1 relative to the benchmark with the status quo as policy outcome and no voting blocs.

Agent r chooses an agenda a_r that maximizes the expected utility given the bloc proposal and acceptance strategies of all other agents. Agenda a_r^1 generates a strictly positive payoff for r. If a_r^1 is not the best response by r, the equilibrium agenda must yield an even greater payoff to r. To do so, it must be an agenda a_r away from the status quo and at least one voting bloc must form (at least with positive probability).

Suppose only bloc $B_1 \subset \mathcal{N}$ forms. If only one voting bloc forms, it cannot contain two agents with ideal policies p_i and $p_j = -p_i$, since they cannot both benefit from the formation of a bloc. Let B_2 be the agents with opposite preferences, formally $B_2 = \{j(i) : i \in B_1\}$, where $j(i)$ is as defined in the proof of proposition 2. Consider any pair of agents (i, j) such that $i \in B_1$ and $j = j(i) \in B_2$. Any policy p^t generates an aggregate loss in utility for i and $j(i)$ unless it lies in the line that connects p_i and p_j. The equilibrium payoffs for r she proposes a_r generate her an expected utility gain of at least β, so these policies are (at least for some issues) away from $(0, 0)$. Since the gains only occur if B_1 manages to enforce commitments, the utility gain for r subject to B_1 enforcing commitments is at least $\beta/(1-\lambda)$. Since $i \in B_1$, it must be that i also

benefits from the policies in a_r that pass if B_1 forms and succeeds in the enforcement of commitments. These policies are either away from the line that connects p_i and $p_{j(i)} = -p_i$, or at the line. If they are all at the line, they generate no aggregate loss or gain for i and j together, but since they benefit r with $x_r > 0$, they are to the right of the midpoint $(0, 0)$ and they unambiguously hurt either i or j; we know they benefit i, hence they must hurt j. If some are away from the connecting line, these generate an aggregate loss for i and j; since the net difference in utility for i must remain positive, j attains a net loss in utility. Let $f(\delta)$ be the minimum possible utility loss for $j = j(i)$ generated by a set of policies that, if implemented, make $i \in B_1$ strictly better off, and make r strictly better off by at least δ. Then, with agenda a_r and a unique voting bloc B_1, if B_1 enforces commitments, j attains a utility loss of at least $f(\beta/(1-\lambda))$. The function $f(\delta)$ is increasing on the gains for the agenda setter and it is strictly positive for all $\delta > 0$. The agents in B_2, by forming a bloc, would prevent the agenda from passing and the utility loss from materializing with probability $(1-\lambda)(1-\lambda)$. With probability $(1-\lambda)\lambda$, B_2 forms and enforces commitments, but its effort is unnecessary because B_1 fails to enforce commitments. With probability λ, bloc B_2 fails to enforce commitments. If c is low enough, $(1-\lambda)(1-\lambda)f(\beta/(1-\lambda)) > c$. But i was arbitrary, so for each $i \in B_1$ and $j(i) \in B$, there exists some $c(p_i)$ such that if $c < c(p_i)$, j would be strictly better off if B_2 forms a second voting bloc. Let $c(B_2) = \min_{i \in B_1} c(p_i)$. If $c < c(B_2)$, every $j \in B_2$ would be strictly better off forming voting bloc B_2. If B_2 does not form, it is a coordination failure. Hence, at least two voting blocs must form in any equilibrium without coordination failures. ■

Proposition 5

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

1) Let $W(A)$ be the set of policies that beat $(0, 0)$ if A forms and $\varepsilon = 0$. Let $W^C(A)$ be its complement. For any $\varepsilon > 0$, let $W_\varepsilon(A)$ be the set of policies that beat $(0, 0)$ if A forms. Let $int(W(A))$ be the interior of $W(A)$. For any point $z = (x_z, y_z)$ and set

V, let $d(z, V)$ be the Euclidean distance between z and V. For any $\delta > 0$ and for any $w \in W(A)$ such that $d(w, W^C(A)) \geq \delta$, $\exists \varepsilon > 0$ such that $w \in W_\varepsilon(A)$. Similarly, for any $\delta > 0$ and any $z \in W^C(A)$ such that $d(w, W^C(A))$, $\exists \varepsilon > 0$ such that $w \in W^C_\varepsilon(A)$, so as $\varepsilon \longrightarrow 0$, the set of policies that can pass given that A forms becomes arbitrarily close to the set of policies 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 A forms given ε converges to the utility of forming A given $\varepsilon = 0$. If no blocs form, with $\varepsilon > 0$, the status quo is not always the policy outcome, but as $\varepsilon \longrightarrow 0$, the set of policies 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 A relative to the benchmark with no blocs converges to the net utility of forming A given $\varepsilon = 0$. For $\varepsilon = 0$ and $c = 0$, members of A attain a strict net gain in utility by forming A. Let β denote the net utility gain made by the member of A who attains the smallest gain if $\varepsilon = 0$ and $c = 0$. For $c = 0$ and ε sufficiently close to zero, members of A still attain a strict net gain in utility if they form A. For an arbitrary $\lambda \in (0, 1)$, let $\overline{\varepsilon} > 0$ be such that if $c = 0$, agents of A attain a net gain in utility of at least $\lambda\beta$ by forming A for any realization of preferences $\widetilde{p} \in N(p, \overline{\varepsilon})$. Then, for any $\varepsilon < \overline{\varepsilon}$, and preferences $\widetilde{p} \in N(p, \varepsilon)$, if $c < \lambda\beta$, agents in A achieve a strictly higher utility forming A than not forming A. The uncovered set converges to the status quo as ε converges to zero, hence given a small enough ε, if A forms policies 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 r can choose an agenda (x, y) on every odd issue and $(x, -y)$ on every even issue such that all agents with $p_i = (1, 0)$ favor it. If c_a is low enough and agents in A form a unique voting bloc, this agenda passes and all the agents in set 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 a_r^1 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 r is strictly better off. It follows that the best agenda a_r also generates a utility gain for r of at least $\beta(\varepsilon)$. The minimum possible utility gain $\beta(\varepsilon)$ is now a function of ε because as ε increases, r may benefit less from the policies in agenda a_r^1. Suppose B_1 forms a unique voting bloc after a_r becomes the agenda. For an arbitrary $i \in B_1$, consider agent $j(i)$ with $p_j = -p_i$. For ε sufficiently small, if a_r benefits r by at least β and it strictly benefits i, it must strictly hurt j. Let $f(x, \varepsilon)$ be the minimum possible loss to j given policy outcomes that make i strictly better off and make r strictly better off by at least x. This function is decreasing in ε, and as $\varepsilon \longrightarrow 0$, it converges to $f(x)$ defined in the proof of proposition 4, so for a small enough ε, $f(x, \varepsilon) > 0$. Then, given a small enough ε, there exists $c_\varepsilon > 0$ such that $(1-\lambda)(1-\lambda)f(\beta(\varepsilon)/(1-\lambda), \varepsilon) > c$ for any $c < c_\varepsilon$. Therefore, given ε small enough, and given $c < c_\varepsilon$, agent $j \in B_2$ would be strictly better off if B_2 forms. Find the lowest such threshold among all members of B_2. If the cost is lower than this threshold, every $j \in B_2$ 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] John H. Aldrich. *Why Parties? The Origin and Transformation of Political Parties in America.* Chicago, University of Chicago Press, 1995.

[2] Scott Ashworth and Ethan Bueno de Mesquita. Informative Party Labels with Institutional and Electoral Variation. *Journal of Theoretical Politics* 20(3): 251—273, 2008.

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

[4] Timothy Besley and Stephen Coate. An economic model of representative democracy. *Quarterly Journal of Economics*, 112(1): 85—114, 1997.

[5] William T. Bianco and Itai Sened. Uncovering evidence of conditional party government. *American Political Science* Review 99(3): 361—371, 2005.

[6] Francis Bloch. *The Endogenous Formation of Economic Coalitions*, chapter Non-Cooperative Models of Coalition Formation in Games with Spillovers. Edward Elgar Publishing, 2003.

[7] Francis Bloch and Armando Gomes. Contracting with externalities an outside option. *Journal of Economic Theory*, 127(1): 171—201, 2006.

[8] Bernard Caillaud and Jean Tirole. Party governance and ideological bias. *European Economic Review,* 34: 779—789, 1999.

[9] Bernard Caillaud and Jean Tirole. Parties as political intermediaries. *Quarterly Journal of Economics*, 117(4): 1453—1489, 2002.

[10] Carlo Carraro. The Endogenous Formation of Economic Coalitions. Edward Elgar Publishing, 2003.

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

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

[13] Gary W. Cox and Matthew D. McCubbins. *Legislative Leviathan.* Berkeley, University of California Press, 1993.

[14] Gary W. Cox and Matthew D. McCubbins. *Legislative Leviathan: Party Government in the House.* Cambridge University Press, 2007.

[15] Daniel Diermeier and Antonio Merlo. Government turnover in parliamentary democracies. *Journal of Economic Theory 94(1): 46—79, 2000.*

[16] Maurice Duverger. *Political Parties.* London, Methuen, 1959.

[17] Jon X. Eguia. Voting blocs, coalitions and parties. Caltech Social Sciences wp 1257, nov 2006.

[18] Justin Fox. Legislative Cooperation among Impatient Legislators. *Journal of Theoretical Politics,* 18(1): 68—97.

[19] Sergiu Hart and Mordechai Kurz. Endogenous formation of coalitions. *Econometrica*, 51(1): 1047—1064, 1983.

[20] Simon Hix, Abdul Noury and Gerard Roland. *Democratic Politics in the European Parliament.* Cambridge University Press, 2007.

[21] Macartan Humphreys. Theory of coalitions and its applications. *Annual Review of Political Science*, 11: 351—386, 2008.

[22] Kyle Hyndman and Debraj Ray. Coalition formation with binding agreements. *Review of Economic Studies*, 74(4): 1125—1147, 2007.

[23] Matthew Jackson and Boaz Moselle. Coalition and party formation in a legislative voting game. *Journal of Economic Theory*, 103(1): 49—87, 2002.

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

[25] Keith Krehbiel. *Pivotal Politics: A Theory or U.S. Lawmaking*. University of Chicago Press, 1998.

[26] Gilat Levy. A model of political parties. *Journal of Economic Theory,* 115(2): 250—277, 2004.

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

[28] Massimo Morelli. Party formation and policy outcomes under different electoral systems. *The Review of Economic Studies*, 71(3): 829—853, 2004.

[29] John Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36: 48—49, 1950.

[30] Martin Osborne and Rabee Tourky. Party formation in single issue politics. *Journal of the European Economic Association,* forthcoming.

[31] Charles R. Plott. A notion of equilibrium and its possibility under majority rule. *American Economic Review*, 57(4): 331—347, 1967.

[32] Debraj Ray. *A Game-Theoretic Perspective on Coalition Formation*. Oxford University Press, 2007.

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

[34] Norman Schofield and Itai Sened. *Multiparty Democracy.* Cambridge University Press, 2006.

[35] Thomas Schwartz. Collective choice, separation of issues and vote trading. *American Political Science Review*, 71(3): 999—1010, 1977.

[36] James Snyder and Michael Ting. An informational rationale for political parties. *American Journal of Political Science*, 46: 90—110, 2002.

[37] Thomas Stratmann. The effects of logrolling on congressional voting. *American Economic Review*, 82(5): 1162—1176, 1992.

[38] Robert M. Thrall and William F Lucas. N-person games in partition form. *Naval Research Logistics Quarterly,* 10: 281—298, 1963.

[39] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior.* Princeton, Princeton University Press, 1944.

[40] Sang-Seung Yi. *The Endogenous Formation of Economic Coalitions*, chapter Endogenous Formation of Economic Coalitions: A Survey of the Partition Function Approach. Edward Elgar Publishing, 2003.