Voting Blocs, Party Discipline and Party Formation.*

Jon X Eguia†
New York University

March 29, 2010

Abstract

I study the strategic incentives to coordinate votes in an assembly. Coalitions form voting blocs, acting as single players and affecting the policy outcome. In an assembly with two exogenous parties I show how the incentives to accept party discipline depend on the types of the agents, the sizes of the parties, and the rules the blocs use to aggregate preferences. In a game of fully endogenous party formation, I find sufficient conditions for the existence of equilibria with one bloc, two blocs, and multiple blocs.

JEL Classification: D71, D72.

Keywords: Voting blocs, party formation, party discipline, coalition formation, voting rule.

*I thank Francis Bloch, Anna Bogomolnaia, Luis Corchon, Matias Iaryczower, Matt Jackson, Patrick LeBihan, Alessandro Lizzeri, Andrea Mattozzi, Tom Palfrey and Andrea Prat for their comments and suggestions.

†Email: eguia@nyu.edu. Mail: 19 West 4th St, 2nd Floor, New York, NY 10012. Phone: 1 212 992 8073.

1 Introduction

Democratic deliberative bodies, such as committees, councils, or legislative assemblies across the world choose policies by means of voting. Members of an assembly can affect the policy outcome chosen by the assembly by forming a voting bloc to coordinate their voting behavior. A voting bloc is a coalition with an internal rule that aggregates the preferences of its members into a single position that the whole coalition then votes for, acting as a single unit in the assembly. From alliances of countries in international relations to political parties in legislative bodies, successful voting blocs influence policy outcomes to the advantage of their members. In national politics, legislators face incentives to coalesce into strong political parties in which every member votes according to the party line. Exercising party discipline to act as a voting bloc, strong parties are more likely to attain the policy outcomes preferred by a majority of party members.

I study the strategic incentives to join voting blocs in an assembly with a finite number of agents who make a binary decision. I explain the formation of these blocs in equilibrium as a function of the heterogeneous preferences of the agents. I contribute to the literature on the endogenous formation of political parties in an assembly, showing that legislators have incentives to coalesce into voting blocs to influence the policy outcome solely for an ideological gain, even in the absence of electoral concerns or a distributive dimension.

In previous formal theories of party formation, Snyder and Ting (2002) describe parties as informative labels; Caillaud and Tirole (1999, 2002) focus on the role of parties as information intermediaries that select high quality 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. Baron (1993) and Jackson and Moselle (2002) note that members of a legislative body have incentives to form parties within the legislature, irrespective of the interaction with the voters, to allocate the pork available for distribution among only a subset of the legislators. Aldrich (1995) explains that US parties serve both to mobilize an electorate in favor of a candidate, and to coordinate a durable majority to reach a stable policy outcome avoiding the cycles created by shifting majorities.

In the first part of the paper I consider an assembly with two exogenous political parties, one on each side of the political spectrum. Any of the theories mentioned above could explain the affiliation of individual agents to these parties, which I take as given and fixed in this section. I study the incentives to form a voting bloc that coordinates the votes of its members enforcing party discipline in each of the two parties.

The formation of these voting blocs requires two conditions. The first is an ex-ante incentive compatibility condition: Agents voluntarily choose to accept party discipline because they find it beneficial in expectation. The second is ex-post incentive compatibility: The players must choose to adhere to the institutions of the voting bloc. I focus on the ex-ante incentive condition, assuming that binding contracts guarantee perfect commitment, so that once voting blocs form, party discipline is enforceable. A possible enforcement device is to request agents to make a costly deposit up front to join a bloc, a deposit that is later returned to members who comply with the party discipline, but it is not returned to those who deviate, making it costly to renege from the commitment to accept party discipline. Alternatively, without up front deposits, if a voting bloc can punish members who vote against the party line at no cost, the threat of punishment suffices to discipline members.[1]

I describe the equilibrium conditions for every member of a party to accept party discipline as a function of the types of the agents, the polarization of the assembly, the sizes of the parties and the internal rule that a party uses to aggregate the preferences of its members. I also describe to what extent the results are robust once we relax the assumption of commitment and we take into account the ex-post incentives to adhere or not to the discipline of a voting bloc.

I show that for some preference profiles a party cannot form a voting bloc that always imposes party discipline on all its members, but it can form a voting bloc that every member wishes to join with laxer party discipline, using an internal voting rule that lets members vote freely when there is substantial disagreement within the party. With respect to polarization of preferences, I find that party discipline becomes increasingly difficult to sustain as the parties become more extreme. In fact, a party of sufficiently extreme agents can only form a voting bloc if it uses a very permissive rule that lets members vote freely as soon as two of them disagree with the party line.

Political parties have incentives to form voting blocs, but voting blocs are not only a consequence of political parties and their sophisticated partisan strategies. Rather, the coordination

[1] In applications, punishments against members are always available: Party members may fail to help a defector in her future career goals, they may ostracize her, etc. The crucial requirement of enforcement devices is that the votes in the assembly be observable. All relevant applications meet this requirement: Voting in democratic legislatures is not only observable by legislators, but in many cases it is in fact publicly available data; voting in many international organizations is also observable.

of votes and the gains to be made by forming a voting bloc are in itself a reason for the endogenous formation of parties.

In the second part of the paper I consider an assembly without pre-existing parties. In this assembly, any subset of voters can coordinate and coalesce to form a voting bloc, and I analyze the endogenous formation of voting blocs in the equilibrium of a game of coalition formation and voting. I find that voting blocs form and exercise party discipline to coordinate the votes of their members. This coordination of votes, under certain conditions, alters the policy outcome. I show that in equilibrium, voting blocs that affect the policy outcome must be of size less than minimal winning. I also find that in a sufficiently polarized assembly, there exists an equilibrium in which exactly two voting blocs form, one on each side of the ideological divide.

In the context of legislative bodies, I interpret these endogenous voting blocs as strong political parties that exercise party discipline. Political parties also play a crucial role in the nomination process and election of their members who run for office, but unlike most of the theories of party formation I have discussed above, my theory does not rely on electoral concerns. In the terminology of Duverger (1959), I study "parties of parliamentary origin," in which a group of legislators forms a party, which only later creates electoral committees. 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 and Calvert and Dietz (2006) introduce externalities, but they all limit their analysis of party formation to assemblies with three agents. More generally, Fox (2006) shows that a single dominant coalition of legislators have incentives to cooperate in an infinitely repeated game. Furthermore, Cox and McCubbins (1993) and Diermeier and Vlaicu (2008) find that legislators in the majority party use the party as means to control the agenda. All these papers deal with the formation of a legislative majority, and cannot explain the formation of two or more parties.

I study assemblies of any size and I find sufficient conditions for equilibria in which two parties form, and in which multiple parties form. The incentives to form parties emerge solely from the heterogeneous preferences over a given policy proposal. While complementary to alternative theories, the rationale for party formation that I propose requires only a proposal put to a vote in an assembly and the capacity to make commitments by the members of the assembly, which makes it more widely applicable than theories that rely on other external actors. For instance, electoral concerns cannot explain the formation of pan-European parties in the European Parliament. European parties, such as the conservative European People's Party or the socialist Party of European Socialist do not run electoral campaigns. Rather, campaigns are run in each State of the EU by its own national parties, such as the German conservative CDU and the socialist SPD, and these national parties, once they win seats in the European Parliament, merge in 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 MEPs have incentives to form stable transnational party organizations and to use these organizations to compete over European Union policies."

I provide a general theory that explains why agents (like-minded or not) in an abstract assembly have incentives to form parties to compete over policies. In a companion paper (Eguia forthcoming) I tailor the model to fit the decision-making process in the United States Supreme Court. In such a restricted setting with only nine agents and simple majority rule, and with an appropriate refinement of the solution concept, I fully characterize the solution set.

The theory applies as well to non-legislative and non-elected assemblies such as international organizations that make policy decisions –or policy recommendations- by voting. Countries such as the members of the EU or the Arab League have incentives to coordinate their votes forming voting blocs in the UN General Assembly or the IMF.

This theory builds upon previous work in the subfield of coalition formation. After the seminal work of Buchanan and Tullock (1962), Hart and Kurz (1983) study the endogenous formation of efficient coalitions. Carraro (2003) and Ray (2007) survey newer theories of coalition formation. In contrast to traditional hedonic models, the *partition function* approach first used by Thrall and Lucas (1963) recognizes that members of a coalition are affected by the actions of agents outside this coalition, and it defines utilities as a function of the whole coalition structure in the society. Bloch (2003) and Yi (2003) survey the literature on coalitions that generate positive externalities or negative externalities to non-members. Bloch and Gomes (2006) propose a general model to cover a variety of applications with either positive or negative externalities. Hyndman and Ray (2007) make the first contribution to the nascent literature on coalitions that generate both positive and negative externalities in a restricted model with only three agents. My theory provides intuitive results for the mixed or hybrid case in which the formation of a voting bloc or party generates both positive and negative externalities to non-members, in a simple framework where the outcome of a voting game determines the payoff to each of finitely many agents.

In the following sections I attempt to apply the game-theoretic insights of the coalition formation literature to shed light on the political economy problem of coordinating the voting behavior of the members of a coalition. First, I present two examples to illustrate how the formation of voting blocs affects voting results and policy outcomes.

1.1 Motivating Examples

After a simplistic illustration of how voting blocs work, I present a more complete example with two voting blocs in a small assembly.

Example 1 *Let there be five agents who must make a binary choice -to approve or reject some action- by simple majority. Suppose that each agent i favors the action with an independent probability $\frac{1}{2}$. Then, the probability that at least three agents favor the action and the action is approved is also $\frac{1}{2}$. The outcome coincides with the preference of agent i if at least two other agents have the same preference as i. This event occurs with probability $\frac{11}{16}$.*

Suppose three agents form a voting bloc with simple majority, so if any two members agree, the third votes with them regardless of her own preference. Then the decision reached depends exclusively on the bloc. The probability that the outcome coincides with the preference of a member i is equal to the probability that at least one other member of the voting bloc has the same preference, which is $\frac{3}{4} > \frac{11}{16}$. Hence, the agents who form a bloc increase the probability that the outcome coincides with their wishes.

Note that a bloc of size two cannot attain a net gain for its members: Either both members agree and vote as they would in the absence of a bloc, or if they disagree, any coordination would result in a loss for one member that fully offsets the gain for the other member. Since there is no gain to be made from coordinating in a bloc of size two, I assume that the minimum size of a bloc that coordinates its votes is three. Then, a bloc of three agents in example 1 is such that no member wants to leave it. However, the bloc with three members is not an equilibrium outcome if outsiders are free to join in. Indeed, both outsiders want to join in.

In example 1, the agents are identical random voters, so that only the size of the bloc matter, not the characteristics of its members. The voting power literature studies this random voting model,[2] calculating the probability that an agent casts a decisive vote. Felsenthal and Machover (2002), and Gelman (2003) among others, analyze how the formation of blocs and alliances affects the probability of casting a decisive vote. I focus instead on the probability of

[2] Hosli (2000, 2002) is an exception that takes into account preferences to calculate the probability that specific sets of agents vote together in the European Council.

getting a desired policy outcome, regardless of the margin of victory. Laruelle and Valenciano (2005) provide a rigorous analysis of the difference between maximizing power, defined as the probability of being able to alter the outcome, and maximizing success, defined as the probability of achieving a desired outcome. The concern for power that guides the voting power literature applies to situations where the goal of the agents is to win by one vote. I am interested in policy decisions where the goal of the agents is to get their desired policy implemented, be it by a marginal vote, or by a landslide.

The random voting model is a non-generic case, in which all agents are ex-ante identical. I am interested in the general case with heterogeneous agents, some of whom are ex-ante more likely than others to favor the action or policy proposal that is put to a vote. The similarity or antagonism of preferences across agents is a key factor in the strategic incentives to coalesce in voting blocs or parties. I assume that there is some uncertainty about preferences, but I assume that ex-ante it is possible to differentiate agents according to their preferences.

We can interpret the uncertainty about preferences in two ways. First, if there is a time difference between the moment when agents coalesce and the time of voting in the assembly, then at the time agents commit to act together, they may not fully know which outcome they will prefer at the time of voting. Three legislators may sign a pact today to vote together in the future, without knowing today the details of the policies they will vote on in the future. Alternatively, in a world of complete information in which agents vote repeatedly, a legislator who votes for the liberal policy with a certain frequency x can be modeled as a legislator with a probability x of voting for the liberal policy in a one-shot voting game.

The ex-ante differences in the preferences of the agents determine the formation of voting blocs. Let us see how a polarized small assembly may split into two different voting blocs, none of which is minimal winning.

Example 2 *Let $\mathcal{N}$ be an assembly with seven agents who must make a binary choice decision -pass or reject some policy proposal- by simple majority. Let each agent i favor the proposal with an independent probability t_i. Suppose $t_1 = t_2 = t_3 = 0.2$, $t_4 = 0.5$ and $t_5 = t_6 = t_7 = 0.8$. Table 1 below shows the probability that the policy proposal passes (column one) and the probabilities that the outcome coincides with the preferences of agents* 1 *(column two),* 4 *(column three) and* 5 *(column four), given that the agents form the voting blocs indicated in each row. If a bloc forms, the whole bloc votes according to the preference of the majority of its members, and in case of a tie, each member votes according to her own preference.*

Table 1: Probabilities that the proposal passes and the agents like the outcome

Bloc	Pass	1	4	5
None	50.0%	61.7%	70.9%	61.7%
{123}	33.3%	67.6%	73.0%	49.2%
{1235}				51.2%
{123},{567}	50.0%	55.1%	90.7%	55.1%
{1234},{567}			67.7%	

The numbers on the table come from simple binomial calculations. The three agents with a low type benefit from forming a voting bloc. However, if other agents can join this bloc, agent 5 benefits from joining. Alternatively, given that the three agents with a low type form a voting bloc, the three agents with a high type benefit from forming their own voting bloc. Consider the outcome with two voting blocs in the fourth row. Since the minimum size for a bloc to coordinate is three, and since the two blocs are symmetric, the table shows that no agent wants to become an independent making her bloc ineffective (second row), and furthermore, even if the political environment is such that agents can switch blocs, no agent wants to switch (third row) and the independent agent four does not want to join either of the two blocs (fifth row). The strategic incentive of agent four to remain independent illustrates an important result: With a purely ideological motivation of caring for the policy outcome and no rents to distribute among the members of the winning coalition, voting blocs in equilibrium are not of minimal winning size. I formalize this intuition in proposition 11.

The insights gained in the previous abstract two examples have important applications to voting in committees, councils, assemblies, and, in particular, in legislatures where legislators can coalesce into political parties that function as voting blocs. Whenever a bloc changes the outcome by casting all its votes according to the preferences of its internal majority instead of splitting its vote according to the preferences of all its members, it benefits a majority of members and hurts only a minority, thus producing a net gain for the bloc as a whole. But generating a gain is not sufficient for the bloc to form. Rather, it must be that every agent has a strategic incentive to participate.

The rest of the paper explores the individual incentives to participate in blocs as a function of the preferences of the members of the assembly and the voting rules used by the voting blocs.

2 Exogenous Parties

Let $\mathcal{N}$ $=\{1, 2, ..., N\}$ be an assembly of voters, where $N \geq 7$ is finite and odd. This assembly must make a binary decision on whether to adopt or reject a policy proposal pitted against a status quo. The division of the assembly is a partition of the assembly into two sets: the set of agents who vote in favor of the proposal, and the set of agents who vote against the proposal. The assembly makes a decision by simple majority and the policy proposal passes if at least $\frac{N+1}{2}$ agents vote in favor.

A voter $i \in \mathcal{N}$ receives utility one if the policy outcome coincides with her preference for or against the proposal and zero otherwise, abstracting away from variation in the intensity of preferences. Let $p_i = 1$ if agent i prefers the proposal to pass, and $p_i = 0$ otherwise; let $p = (p_1, ..., p_N)$ be a preference profile for the whole set of voters, and let $p_{-i} = (p_1, ..., p_{i-1}, p_{i+1}, ..., p_N)$ be the profile without the preference of i. Similarly, let $v_i = 1$ if agent i votes in favor of the proposal in the division of the assembly, and $v_i = 0$ otherwise, and let $v = (v_1, ..., v_N)$.

Agents face uncertainty at the initial stage. They do not know the profile of preferences in favor or against the proposal. They only know, for each profile of preferences $p \in \{0, 1\}^N$, the probability that p occurs. Let $\Omega : \{0, 1\}^N \longrightarrow [0, 1]$ be the probability distribution over profiles and assume Ω is common knowledge. Let the type of agent i be the probability that i favors the policy proposal. Note that this is an unconventional use of the word *type:* In this model, types refer to probabilities that are common knowledge. If these probabilities are not correlated across agents, then I say that types are independent. Let $P[\cdot]$ denote the probability of an event.

Definition 1 *Let $t_i \equiv P[p_i = 1]$ be the type of agent i. Types are independent if $P[p_i = 1|p_{-i}] = t_i$ for all $i \in \mathcal{N}$ and all $p_{-i} \in \{0, 1\}^{N-1}$.*

Types are independent if the probability that i favors the policy proposal is t_i for any given realization of preferences by the rest of agents. Equivalently, we may say that types are independent if $\Omega(p) = \prod\limits_{i\in\mathcal{N}} [t_i p_i + (1 - t_i)(1 - p_i)]$ for any $p \in \{0, 1\}^N$.

Let the assembly be composed of two exogenously given coalitions L and R, which I call "parties" and a set M of independent agents who belong to neither of the two parties, so $\mathcal{N}$ $=L \sqcup R \sqcup M$, where $\sqcup$ denotes the union of two disjoint sets. Let N_L, N_R and N_M be the respective sizes of L, R and M and assume for simplicity that all three sizes are odd. Each party can enforce party discipline to coordinate the voting behavior of its members by forming a voting bloc. A voting bloc is defined as a set of agents and an internal aggregation rule to aggregate preferences into votes. For a party $J \in \{L, R\}$ and for any $C \subseteq J$, let p_C and v_C denote restrictions of the preference profile p and the voting profile v to the members of subset C of party J. Let $\hat{p}_i \in \{0,1\}$ be the preference that $i \in L \sqcup R$ reports to its party, and let $\hat{p}_C$ be the reported preferences of all members of C.

An internal aggregation rule $r_C(\hat{p}_C) : \{0,1\}^{N_C} \longrightarrow \{0,1\}^{N_C}$ is a mapping from the reported preferences to the votes that C casts in the assembly, so $r_C(\hat{p}_C) = v_C$. For party J with $J \in \{L, R\}$, and any set $C \subseteq J$, I focus on aggregation rules that are quota rules (anonymous, neutral, decisive and monotonic aggregation rules) characterized by a scalar $r_J \in \left(\frac{1}{2}, 1\right]$ with the interpretation that for a set of agents C, if at least a fraction r_J of the members of C prefer (oppose) the policy proposal, then everyone in C casts a vote for (against) the proposal. If the fractions of members in favor and members opposed are both below r_J, then each member votes according to her own preference. Given a set of agents $C \subseteq J$, an internal rule $r_C(\hat{p}_C)$ characterized by scalar r_J is then:

$$r_C\left(\hat{p}_C\right) = \left\{ \begin{array}{c} (0,0,...,0) \ \ if \ \sum\limits_{i \in C} \hat{p}_i \leq (1 - r_J)N_C, \\ \hat{p}_C \ \ if \ (1 - r_J)N_C < \sum\limits_{i \in C} \hat{p}_i < r_J N_C, \\ (1,1,...,1) \ \ if \ \sum\limits_{i \in C} \hat{p}_i \geq r_J N_C. \end{array} \right\}$$

Let the pair (C, r_J) denote an arbitrary voting bloc, where $C \subseteq J$ is the set of members of party J who accept party discipline and form the bloc, and r_J is the scalar that characterizes the voting rule used by the bloc. I assume that each member of a party voluntarily chooses whether or not to accept party discipline and join her party's voting bloc. For any $i \in L \sqcup R$, let $a_i \in \{0,1\}$ denote the decision by agent i, where $a_i = 1$ if agent i accepts party discipline and joins her party's voting bloc. Agents who choose $a_i = 0$ reject party discipline, do not join the voting bloc, and do not coordinate their votes, behaving as if they were independents. Independent agents are not affiliated to a party and do not coordinate their votes.

The timing of events is as follows.

1. Given two scalars $r_L \in \left(\frac{1}{2}, 1\right]$ and $r_R \in \left(\frac{1}{2}, 1\right]$, each party member i simultaneously chooses a_i. Voting blocs (C_L, r_L) and (C_R, r_R) form, where $C_J = \{i \in J : a_i = 1\}$ for $J \in \{L, R\}$.

2. A preference profile p is realized and each agent i privately learns her own preference p_i.

3. Voting blocs (C_J, r_J) meet, every $i \in C_J$ reports $\hat{p}_i$ and C_j aggregates its internal preferences according to its internal rule r_J. The outcome of these meetings becomes common knowledge.

4. The whole assembly meets. The aggregation at step 3 is binding and members of blocs

vote according to the outcome of their bloc's internal meeting. Each agent i who joined no bloc chooses $v_i \in \{0,1\}$.

I refer to "a scalar that characterizes an internal aggregation rule" simply as "a rule". Given an internal rule r_J, and a set of agents $C \subseteq J$, the majority inside voting bloc (C, r_J) must gather at least $r_J N_C$ votes in order to exercise party discipline, reverse the votes of the internal minority and make C act as a unitary player in the assembly. A rule with $r_J = \frac{N_C+1}{2N_C}$ is simple majority, and $r_J = 1$ is unanimity, which is identical to not coordinating any votes –members only vote together if they all share the same preference. Note that if $r_J N_C$ is not an integer, the threshold of majority votes necessary for the bloc to reverse the votes of the internal minority is $\lceil r_J N_C \rceil$, where $\lceil x \rceil \equiv \min\{k \in \mathbb{N} : k \geq x\}$ is the ceiling function. The largest internal minority whose votes are ever reversed is of size $N_C - \lceil r_J N_C \rceil = \lfloor (1 - r_J) N_C \rfloor$, where $\lfloor x \rfloor \equiv \max\{k \in \mathbb{Z} : k \leq x\}$ is the floor function.

Members of a voting bloc reveal their private preference by voting in the internal meeting of the bloc. Since there are only two alternatives, and the rules of both blocs and the assembly are such that the probability that each alternative wins is increasing in the number of votes it receives, there is no profitable deviation from sincere voting; voting against her preference can only make an agent worse off. Note that sincere voting for agents who accept party discipline and join a voting bloc consists of voting according to their preference at the internal meeting of their voting bloc. In the assembly these agents no longer have a strategic choice to make, and their vote is cast mechanically by their voting bloc as determined during the internal meeting.

Lemma 1 *Given any agent $i \in \mathcal{N}$, and any actions at the bloc formation stage by every agent in $L \sqcup R$, for any $i \in M$ and for any $i \in L \sqcup R$ s.t. $a_i = 0$, $v_i = p_i$ is a weakly dominant strategy at the voting stage. For any $i \in L \sqcup R$ such that $a_i = 1$, sincerely voting $\hat{p}_i = p_i$ is a best response to any strategy profile by agents in $\mathcal{N} \backslash \{i\}$ such that $\{j \in M \text{ or } a_j = 0\} \Longrightarrow \{v_j = p_j\}$.*

I prove lemma 1 and all other results in the appendix. Based on lemma 1, I assume in what follows that members of a bloc reveal their preference truthfully, so $\widehat{p}_i = p_i$ for all $i \in C$, all $C \subseteq \mathcal{N}$ and all $p \in \{0,1\}^N$. In a more straightforward interpretation of the model that bypasses internal voting as a strategic choice, a bloc learns the true preferences of its members, and its aggregation rule maps the internal preferences into a number of votes to be cast in favor of the policy proposal in the division of the assembly.

Given the sizes of the parties, the distribution of types, and the internal voting rules for each party, agents play a coalition formation and voting game. Since voting sincerely is always a best response to any undominated strategy by every other agent, I focus only on the coalition-formation game played by every agent who belongs to party L or party R. Agents maximize the ex-ante probability that the policy outcome in the assembly coincides with their policy preference, by strategically choosing whether or not to participate in their party's voting bloc, accepting party discipline. The expected payoffs are determined by the probability over outcomes given by the voting blocs that form, the rules of each party, and the anticipated sincere votes; the expectation is over the realization of preferences given the distribution of types.

I seek to explain the development of party discipline inside two strong parties in which all members coordinate their votes and the party acts as a unitary actor in the assembly. The equilibrium concept I use is Nash equilibrium in pure strategies and I find necessary and sufficient conditions for an equilibrium to exist in which every party member accepts party discipline.

I initially assume that binding contracts guarantee perfect commitment, so that once blocs form, ex-post agents cannot renege from their commitment to vote as dictated by the internal rule of the bloc. I relax this assumption at the end of the section, introducing ex-post incentive conditions to adhere to party discipline at the voting stage in the assembly.

To capture the insight that party membership is correlated with policy preferences, I assume that party L leans left and tends to vote in the aggregate against the policy proposal, while party R leans right and with high probability a majority of its members favor the policy proposal. This assumption allows for widespread overlap on the types of members of the two parties: Party L may have one or several high types in its ranks; all I require to say that L leans left is that for any $q > 0$, if L has k members with types above $\frac{1}{2} + q$, then L has at least k members with type below $\frac{1}{2} - q$; formally,

$$|\{i \in L : t_i \leq x\}| \geq |\{i \in L : t_i \geq 1 - x\}| \ \forall x \in [0, 1/2].$$

Here is an equivalent definition.

Definition 2 *Let $C \subseteq \mathcal{N}$ of size N_C be a set of agents with independent types, and let the agents be labeled from 1 to N_C so that $i \leq j \iff t_i \leq t_j$. We say C leans left if $t_i + t_{N_C+1-i} \leq 1$ for any $i \in \{1, 2, ..., N_C\}$. Coalition C leans right if the inequality is reversed, and is symmetric if the condition holds with equality.*[3]

In a left leaning party there may be several agents with high types or even some with very high types, but there are more agents with low types and very low types, respectively.

[3] All the results in this section hold for a weaker but more cumbersome condition, available from the author.

Specifically, the $n-th$ lowest type in a left leaning party C is weakly closer to zero (more leftist) than the $n-th$ highest type is to one, for any n. As suggestive evidence that the restriction is mild and satisfied in practice, consider the American Conservative Union scores, which measure the frequency with which a legislator votes according to the wishes of the Union). Letting t_i be the score of each US senator in 2008, L be the Democrats and R the Republicans, $t_i + t_{N_L+1-i} \in [0.08, 0.32]$ $\forall i \in \{1, ...N_L\}$, while $t_j + t_{N_R+1-j} \in [1.12, 1.55]$ $\forall j \in \{1, ...N_R\}$.[4]

Assuming that one party leans left, the second party leans right, and the set of independent agents is symmetric, I find the necessary and sufficient condition on the types of the members of each party in order for every party member to join their voting bloc in equilibrium.

For any $i \in J$, with $J \in \{L, R\}$, let π_i^{1+} and π_i^{1-} respectively be the probability that i is better off (π_i^{1+}) and worse off (π_i^{1-}) if i accepts party discipline and joins bloc (J, r_J) than if she does not, subject to $p_i = 1$, and subject to every other member of J accepting party discipline, and similarly let π_i^{0+} and π_i^{0-} respectively be the probability that i is better off (π_i^{0+}) and worse off (π_i^{0-}) if i joins bloc J subject to $p_i = 0$ and every other party member accepting discipline. Then, the net benefit for agent i of accepting party discipline assuming that every other agent in J does so is

$$t_i(\pi_i^{1+} - \pi_i^{1-}) + (1 - t_i)(\pi_i^{0+} - \pi_i^{0-}).$$

For any $i \in J$, with $J \in \{L, R\}$, let

$$P_i^{1+} \equiv P\left[\sum_{i \in J \setminus \{i\}} p_i = \lceil r_J N_J \rceil - 1\right] P\left[\sum_{m \in \mathcal{N} \setminus J} v_m \in [\frac{N+1}{2} - N_J, \frac{N-1}{2} - \lceil r_J N_J \rceil]\right],$$

$$P_i^{1-} \equiv P\left[\sum_{i \in J \setminus \{i\}} p_i \leq \lfloor (1 - r_J) N_J \rfloor - 1\right] P\left[\sum_{m \in \mathcal{N} \setminus J} v_m = \frac{N-1}{2}\right],$$

$$P_i^{0+} \equiv P\left[\sum_{i \in J \setminus \{i\}} p_i = \lfloor (1 - r_J) N_J \rfloor\right] P\left[\sum_{m \in \mathcal{N} \setminus J} v_m \in [\frac{N+1}{2} - \lfloor (1 - r_J) N_J \rfloor, \frac{N-1}{2}\right],$$

$$\text{and } P_i^{0-} \equiv P\left[\sum_{i \in J \setminus \{i\}} p_i \geq \lceil r_J N_J \rceil\right] P\left[\sum_{m \in \mathcal{N} \setminus J} v_m = \frac{N+1}{2} - N_J\right].$$

If $\lceil r_J(N_J - 1) \rceil = \lceil r_J N_J \rceil$, then $\pi_i^{1+} = P_i^{1-}$, $\pi_i^{1-} = P_i^{1-}$, $\pi_i^{0+} = P_i^{0+}$, and $\pi_i^{0-} = P_i^{0-}$. Condition $\lceil r_J(N_J - 1) \rceil = \lceil r_J N_J \rceil$ means that the size of the majority necessary to enforce party discipline stays constant if the size of bloc shrinks by one. The events in which an

[4] Using the scores of the American Democratic Action, $t_i + t_{N_L+1-i} \in [0.30, 0.80]$, where L in this case are the Republicans, who are typically against policies favored by the ADA, while $t_i + t_{N_R+1-i} \in [1.45, 1.90]$, where R are the Democrats.

agent strictly prefers to be part of the bloc are then those in which the agent is essential to bring the majority above the threshold that triggers the bloc to coordinate; if the agent is not part of the bloc, the majority falls one short. Rules such that $r_J N_J$ is an integer satisfy this condition and simplify notation, and I use them in my examples. More generally, rules such that $r_J \in (\frac{k-1}{N_J-1}, \frac{k}{N_J}]$ for some $k \in \{\frac{N_J+1}{2}, \frac{N_J+3}{2}, ..., N_J\}$ satisfy the condition.

For any $J \in \{L, R\}$, let $l_J, h_J \in J$ be such that $t_{l_J} \leq t_i \leq t_{h_J}$ for any $i \in J$. These are the agents with the lowest (l) and highest (h) type in each party.

Proposition 2 *Let $\mathcal{N} = L \sqcup M \sqcup R$. Suppose that types are independent, L leans left, M is symmetric, R leans right and $\lceil r_J(N_J-1)\rceil = \lceil r_J N_J \rceil$ for $J \in \{L, R\}$. Then it is a Nash equilibrium for every agent in L and R to respectively join (L, r_L) and (R, r_R) if and only if*

$$t_{h_L} \leq \frac{P^{0-}_{h_L} - P^{0+}_{h_L}}{P^{1+}_{h_L} - P^{1-}_{h_L} + P^{0-}_{h_L} - P^{0+}_{h_L}} \text{ and } t_{l_R} \geq \frac{P^{0-}_{l_R} - P^{0+}_{l_R}}{P^{1+}_{l_R} - P^{1-}_{l_R} + P^{0-}_{l_R} - P^{0+}_{l_R}}.$$

Each party forms a voting bloc with all its members in equilibrium if the highest type in L is not too high, and the lowest type in R is not too low. The types of the members of each bloc may overlap, i.e., the right-most member of the Left party may be to the right of the left-most member of the Right party, but it cannot be too far to the right, and similarly agents too far to the left would not choose to join the Right bloc.

The intuition of the proof takes two steps. First, I show that in each party, there is one well-defined agent who benefits least from joining the voting bloc. If this agent who benefits least wants to join, every other party member joins as well. Intuitively, in party L, dominated by low types, the outcome of the internal vote is often negative, against the policy proposal. Thus the agent with the highest type inside the bloc is the agent who most often loses the internal vote and who benefits least from the coordination of votes. It suffices to check that this agent wants to join the bloc in party L, and that its counterpart with the lowest type in party R wants to join her own bloc to guarantee that both parties form voting blocs with full membership and coordinate the votes of all their members. The second step consists of showing that the extreme member of L who is least willing to join the voting bloc, joins the bloc if her type is below some cutoff function, and, similarly, the member of R least willing to join the voting bloc (R, r_R) prefers to join if her type is high enough given the type of her fellow party members.

If $\lceil r_J(N_J-1)\rceil \neq \lceil r_J N_J \rceil$, then $\lceil r_J(N_J-1)\rceil = \lceil r_J N_J \rceil - 1$ and the incentives to be part of the bloc are less intuitive: An agent strictly prefers to be part of the bloc in events when her participation is essential to *prevent* the bloc from coordinating the votes of its members.

If $\lceil r_J(N_J-1)\rceil = \lceil r_J N_J\rceil - 1$, a majority remains above the reduced threshold after one of its members abandons the bloc. It is the size of the minority necessary to stop the bloc from coordinating votes that stays constant, so the incentives to join the bloc are not to make the bloc work as a coordination device, but to stop it from working as such. I provide the values of π_i^{1+}, π_i^{1-}, π_i^{0+} and π_i^{0-} for this second case in the appendix, as part of the proof of proposition 3.

While proposition 2 shows that the highest type in L must be not too high, and the lowest type in R must be not too low for the two parties to form voting blocs, this result does not imply that extreme parties where all the types are very low or very high coordinate votes more easily. On the contrary: I show that in a sufficiently extreme party, party members reject party discipline under almost all supermajority rules.

Proposition 3 *Let $\mathcal{N} = L \sqcup M \sqcup R$. Suppose $N_L \leq \frac{N-1}{2}$, $N_R \leq \frac{N-1}{2}$, types are independent, M is symmetric, R leans right and forms a voting bloc (R, r_R) and $r_L \leq \frac{N_L-2}{N_L-1}$. There exists $\varepsilon > 0$ such that if $t_{h_L} \leq \varepsilon$, it is not a best response for every $i \in L$ to accept party discipline.*

A corollary to proposition 3 is that no extreme party of size more than three but less than minimal winning can form a voting bloc with simple majority in a pure strategy equilibrium, even if its members all share a common type. The intuition for this negative result for extreme parties is that the preference of the internal majority is all but certain: In an extreme left party, the majority rejects the policy proposal with probability very close to one. In the –almost complete- absence of uncertainty about the result of the internal vote, agents prefer to step out of the voting bloc to avoid voting against their preference when they happen to favor the policy proposal.

The maximum size of parties that can coordinate the votes of their members is inversely related to the extremism of its members. Specifically, I find that in a symmetric assembly with two homogeneous parties of the same size and a set of homogeneous independent agents, the maximum size of the parties such that they can form voting blocs with simple majority is decreasing in the polarization of the two parties.

Claim 4 *Let $\mathcal{N} = L \sqcup M \sqcup R$. Assume $N_L = N_R = n$ odd. Assume $t_i = t_L = \frac{1}{2} - q$ for every $i \in L$, $t_i = \frac{1}{2}$ for every $i \in M$ and $t_i = t_R = \frac{1}{2} + q$ for some $q \in (0, 0.5)$. For any odd $N \leq 325$, there exists a non increasing function $n(q) : (0, 0.5) \longrightarrow \mathbb{N}$ such that an equilibrium in which L and R form voting blocs with simple majority exists if and only if $n \leq n(q)$. Alternatively, for any odd $N \leq 325$, there exists a strictly decreasing function $q(n) : \mathbb{N} \longrightarrow [-1, 0.5]$ such that* *an equilibrium in which L and R form voting blocs with simple majority exists if and only if* $q \leq q(n_L)$.

I find this result computationally, calculating the threshold $\frac{P_i^{0-}-P_i^{0+}}{P_i^{1+}-P_i^{1-}+P_i^{0-}-P_i^{1+}}$ as a function of the polarization q separately for each pair (N, n) such that $N \in [7, 325]$, $n \in [7, \frac{N-1}{2}]$, and N and n are odd.[5] The following example details the case of an assembly of size $N = 101$, considering as well an asymmetric distribution of types so that one party is larger than the other.

Example 3 *Let* $\mathcal{N} = L \sqcup M \sqcup R$*. Suppose* $N = 101$*, types are independent,* $t_i = 1/2$ *for every agent* $i \in M$*,* $t_j = t_L$ *for every member* $j \in L$ *and* $t_k = t_R$ *for every member* $k \in R$*. Columns two and three of the following table show the maximum size of the two parties* L, R *such that voting blocs* (L, r_L) *and* (R, r_R) *form in equilibrium with* $r_C = \frac{N_C+1}{2N_C}$ *for* $C = L, R$*, for a symmetric assembly where* $N_L = N_R$ *(column two) and an asymmetric assembly where* $N_L = 2N_R - 1$ *(column three), given the degrees of polarization specified in the different rows.*

(t_L, t_R)	$N_L = N_R$	$N_L = 2N_R - 1$
$(0.45, 0.55)$	*31, 31*	*29, 15*
$(0.4, 0.6)$	*23, 23*	*25, 13*
$(0.3, 0.7)$	*13, 13*	*17, 9*
$(0.2, 0.8)$	*7, 7*	*9, 5*
$(0.1, 0.9)$	*3, 3*	*5, 3*

The intuition is that extremists are only able to coordinate in small numbers, while moderate agents can form larger voting blocs. If parties are larger than the sizes indicated in the table, their member cannot agree in equilibrium to coordinate their votes under simple majority rule.

Simple majority, which corresponds to any $r_J \in \left(\frac{1}{2}, \frac{N_J+1}{2N_J}\right]$, is the internal rule that maximizes the sum of utilities of the members of a voting bloc $V_C = (C, r_J)$ for any $C \subseteq J$. A voting bloc subtracts votes from its internal minority, adding them to the internal majority. Hence, if the bloc alters the outcome in the assembly, it changes it from the outcome preferred by a minority of the bloc to the one preferred by a majority. Since there is no intensity of preferences, it follows that the sum of utilities in the bloc increases. Simple majority maximizes

[5] Size $N \leq 325$ covers most of the cases of interest, including any international assembly that assigns one vote per country, 18 out of the 32 legislative assemblies of the 17 democratic nations in the $G20$, and all lower and upper legislative assemblies of all 50 US States except the lower house in New Hampshire. The result also holds for all other larger sizes that I checked, including the size of the US House of Representatives $N = 435$. The Mathematica file containing these calculations is available from the author.

the probability that the bloc alters the outcome in the division of the assembly and gains a surplus, and it maximizes the sum of utilities of the bloc.

In some instances, party J members do not all accept party discipline under simple majority, but they all accept party discipline with some other supermajority r_J internal rule, even though the supermajority yields a smaller surplus of aggregate utility for the bloc. As a stark illustrative case, assume that all members of party J share the same type so they benefit equally if J enforces party discipline. Suppose that some party members benefit more by rejecting party discipline, letting others form the bloc and free-riding if the internal voting rule is simple majority. Party J can instead form a voting bloc with a supermajority rule $\frac{N_J-1}{N_J}$. With this rule, if the bloc loses a single member, its internal majority only reaches the coordination threshold if all agents agree, so the bloc never affects the outcome and generates no surplus for anybody. A stringent supermajority rule that makes every agent essential for the bloc to function thus deters defections and free-riding.

Proposition 5 states a slightly more general result, noting that the intuition holds if types in J are similar but not necessarily identical.

Proposition 5 *Let $\mathcal{N} = L \sqcup M \sqcup R$. Suppose that types are independent with $t_i \in (0,1)$ for all $i \in \mathcal{N}$. Suppose that L leans left and forms a voting bloc (L, r_L), M is symmetric, $N_L \leq \frac{N-1}{2}$ and $3 < N_R \leq \frac{N+1}{2}$. There exist $t_R \in (\frac{1}{2}, 1)$, $\varepsilon > 0$ and $\hat{r}_R \in [\frac{N_R+3}{2}, N_R - 1]$ such that if $t_i \in (t_R - \varepsilon, t_R + \varepsilon)$ $\forall i \in R$, it is not a best response for every $i \in R$ to join (R, r_R) if r_R is simple majority, but it is a best response for every $i \in R$ to join $(R, \hat{r}_R)$.*

Some parties that cannot get all their members to join a voting bloc with simple majority, can get all their members to coordinate when the internal majority in the party is more substantial than a mere majority of one. Figure 1 illustrates this result. I plot the lowest value of $t_{l_R} \equiv t_l$ for which R can form a voting bloc as a function of t_R for $r_R = \frac{5}{9}$ (simple majority), $\frac{6}{9}$ and $\frac{8}{9}$. To be able to plot t_l as a function of a single parameter, I let $N_L = 11$, $N_R = 9$, $N_M = 31$, $t_i = 0.3$ for all $i \in L$,[6] $r_L = \frac{6}{11}$ (simple majority), $t_m = 0.5$ for all $m \in M$ and $t_j = t_R$ for all $j \in R \backslash \{l_R\}$.

Observe that in this example, the more stringent the internal voting rule of R, the lower the type of l_R can be such that l_R wants to participate in the voting bloc (R, r_R). This observation does not generalize to all cases. In other examples (available from the author) with heterogenous types within each coalition, a bloc can form in equilibrium with simple majority, but not with a

[6] The assumption that the L members have a common type 0.3 is arbitrary, and a very similar graph results for any vector of types in coalition L such that L votes *no* with probability close to one.

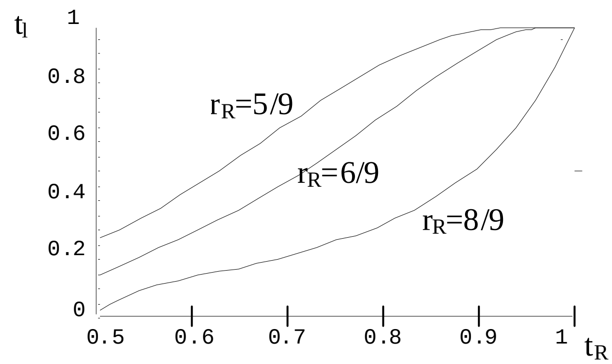

Figure 1: The relevance of the internal rule on the decision to accept party discipline.

supermajority rule. However, since the internal voting rule that maximizes the sum of utilities of the members of a voting bloc among the class of $r - majority$ rules subject to the constraint that every party member joins the bloc is the lowest possible rule satisfying the constraint, the inability to use higher supermajority rules is not as problematic as the inability to use simple majority. Proposition 5 shows that for some parameters, the second best rule that maximizes surplus subject to being accepted by all members is a supermajority.

This result contrasts with the findings of Maggi and Morelli (2006) who study a single coalition that votes on whether or not to take a collective action. They find that the optimal rule in an infinitely repeated game is always either the rule that maximizes the social welfare if agents are patient enough, or unanimity if agents are impatient, and never an intermediate rule. They restrict attention to homogeneous agents (or in their terminology, "symmetric" agents), and in their model the collective action of the coalition does not generate an externality to non-members. Once we take into account that agents are heterogeneous and that the actions of a coalition generate externalities to non-members, I show that a supermajority rule that is not welfare-maximizing for the coalition sometimes becomes the optimal internal rule given the constraint that agents cannot be forced to participate in the collective action -in my case, the coordination of votes- undertaken by the coalition.

This formal result is consistent with the "conditional party government" applied theory of Rohde (1991) and Aldrich and Rohde (2001), who look at party discipline in the US Congress and conclude that back-benchers delegate authority to their leaders to impose a party line only when there is little disagreement within the party. In the words of Cox and McCubbins (1993), page 155:

> The gist of conditional party government is that the party leadership is active

> only when there is substantial agreement among the rank and file on policy goals. If this hypothesis is true, one would expect that decreases in party homogeneity should lead, not to decreases in support given to the leaders when they take a stand, but rather to leaders taking fewer stands. This is essentially what we find.

Proposition 5 shows that this finding is not an idiosyncrasy of the Democratic and Republican parties in the US Congress, but rather, a general principle at work: Party leaders find it easier to make their party work as a disciplined voting bloc if they only enunciate a party line when the minority of dissenters inside the party is small, and they let members vote freely whenever the internal minority is large.

Sometimes a party cannot form a voting bloc under any supermajority rule because it faces a free-riding problem. Every party member would be better off if the party forms a voting bloc, but some individual party members benefit even more if the bloc is formed without them, so they let others coordinate their votes alone. A party can solve this collective action problem if it can commit to coordinate votes only if every party member accepts discipline. Each agent must then weigh the gain brought by the bloc, and not the marginal advantage of being in or out of a bloc that forms.

More generally, expand the theory to allow for a larger class of internal voting rules, including non-anonymous and non-deterministic rules. For any finite set X, let ΔX be a probability distribution over X. Suppose that for $J \in \{L, R\}$, rule $r_J : 2^J \times 2^{C_J} \longrightarrow \Delta\left(\{0,1\}^{N_{C_J}}\right)$, where $C_J = \{i \in J : a_i = 1\}$. The rule now takes two inputs: The subset C_J of party members who accept party discipline and join the bloc, and the subset of bloc members who favor the proposal. The rule maps these inputs into a probability distribution over vectors of votes in the assembly for the members of the bloc. For instance, a rule $r_J(\cdot)$ might dictate that the bloc uses simple majority if every party member joins the bloc, but if agent i fails to join, then the bloc coordinates votes only with probability a half. While non-deterministic rules may be hard to implement in practice, they are theoretically appealing, as shown in the following result, which does not require type independence; it suffices that any realization of preferences occurs with positive probability, and that the rules satisfy a very mild Pareto requirement.

Axiom 6 *Given any set C, an internal voting rule r_C is weakly Pareto optimal if* $\{p_i = k \ \forall i \in C\} \Longrightarrow r_C(p_C) = (k, k, ..., k) \ \forall k \in \{0, 1\}$.

In order to be weakly Pareto optimal, if all agents in C have the same preference, the internal voting rule must let them all vote for their commonly preferred outcome in the assembly.

Proposition 7 *Let* $\mathcal{N} = L \sqcup M \sqcup R$. *Suppose that* $\Omega(p) > 0$ *for any* p, r_L *is weakly Pareto optimal for any* $C_L \subseteq L$ *and* $N_L, N_R \leq \frac{N-1}{2}$. *There exists a rule* $r_R : 2^J \times 2^{C_R} \longrightarrow \Delta\left(\{0,1\}^{C_R}\right)$, *where* $C_R = \{i \in R : a_i = 1\}$, *such that in equilibrium every* $i \in R$ *is strictly better off accepting party discipline under rule* $r_R(\cdot)$.

Propositions 2 and 3 and claim 4 show how the ability of a party to coordinate the votes of its members depends on their types and their ex-ante preferences. The general intuition is that it is easier to coordinate agents with homogeneous and moderate types. Proposition 5 shows that some parties that are unable to always exercise party discipline can exercise party discipline sometimes, forming a voting bloc with a supermajority rule. Proposition 7 goes one step further: Devising more complicated internal rules that call for party discipline to be exercised only in some events, and then only with some probability in those events, a party can guarantee that every member strictly prefers to accept this selective party discipline, regardless of the ex-ante distribution of party members' preferences.

2.1 Extension: No Commitment

The theory in this paper focuses on the ex-ante incentives to coalesce into a voting bloc assuming that commitment is possible.[7] In this subsection I relax the assumption of commitment and I show that parties can still form voting blocs and gain a utility surplus from the coordination of votes.

Assume that voting blocs merely issue voting recommendations. The vote v_i in the assembly is now a strategic choice for each agent. Members of a voting bloc can follow the recommendation, or ignore it. Assume that if they ignore it, they suffer a small penalty or cost $c > 0$. This cost can be arbitrarily small, reflecting a negligible capacity of parties to create incentives to follow their recommendations.

The best response of a voting bloc member in the assembly is to vote $v_i = p_i$ if her vote is pivotal, incurring cost c, but attaining the desired policy outcome; and to follow the recommendation and avoid the cost if her vote is not pivotal. Suppose that parties, anticipating this behavior, can use contingent recommendation rules that request coordination only when no individual agent is pivotal. I distinguish between the voting rule of the bloc, and the recommendation that the rule be followed or not by the bloc. Formally, for each $J \in \{L, R\}$, and any $C_J \subseteq J$, let $r_J \in \left(\frac{1}{2}, 1\right]$ be the scalar that characterizes the internal voting rule of the voting bloc (C_J, r_J), as in the benchmark theory at the beginning of the section. Let $r_J^i(p_{C_J}) \in \{0,1\}$

[7] For an extensive discussion of the challenge of modelling commitment technologies, see the recent survey on coalitions by Humphreys (2008).

be the vote that $i \in C_J$ must cast in the assembly according to rule $r_J(p_{C_J})$. For each party $J \in \{L, R\}$, let $d_J : \left(\frac{1}{2}, 1\right]^2 \times 2^L \times 2^R \times 2^{\mathcal{N}} \longrightarrow \{0, 1\}$ be the recommendation function of party J. Its arguments are the scalars r_L and r_R, the decisions to accept or reject party discipline by each member of L and R, and the realized preference profile p. Given an observed value for each of these variables, the recommendation function recommends either $(d_J = 1)$ that bloc members vote according to the internal rule $v_i = r^i_J(p_{C_J})$, or $(d_J = 0)$ that the members ignore the rule and vote sincerely $v_i = p_i$ in the assembly.

A voting bloc is now defined by its membership, voting rule and recommendation function. The key insight is that with any $r-majority$ rule, including simple majority, and a recommendation function that recommends agents to vote sincerely whenever following the rule would lead to a marginal outcome decided by a single vote, every party member strictly prefers to accept party discipline and follow the recommendations.

Proposition 8 *Let $\mathcal{N} = L \sqcup M \sqcup R$. Suppose that $\Omega(p) > 0$, $N \geq 9$, and for $J \in \{L, R\}$, $3 \leq N_J \leq \frac{N-1}{2}$, $r_J \leq \frac{N_J - 1}{N_J}$ and the recommendation rule d_J is such that $d_J = 1$ if and only if*

$$\sum_{i \in L} r^i_L(p_L) + \sum_{i \in R} r^i_R(p_R) + \sum_{i \in M} p_i \notin \left\{ \frac{N-1}{2}, \frac{N+1}{2} \right\}.$$

Then there exists an equilibrium in which every $i \in L \sqcup R$ joins her party's voting bloc and is strictly better off joining than not joining.

An alternative interpretation of this result is that if there is a party leader or party whip, who aims to coordinate votes to increase the aggregate welfare of party members but lacks enforcement powers, this powerless leader can persuade party members to accept party discipline and follow party voting recommendations by conceding that it is impossible to coordinate votes when the outcome is marginally decided by one vote. The party still profits from voting coordination by changing the outcome in events in which two or more agents vote against their preference.

Notice that informational assumption that the preference profile p becomes known before the vote in the assembly. A party (or party leader or whip) who aims to coordinate its members and lacks enforcement power needs instead good information to anticipate the votes of other legislators; a party whip with neither power nor information is ineffective. For a detailed case study on vote coordination with very weak enforcement, and the crucial importance of information in this environment, see the historical account by Caro (2002) of Lyndon Johnson's term in the US Senate.

3 Endogenous Parties

In this section I fully endogenize parties. The new primitive is an assembly of individual voters who coalesce freely with each other according to their individual strategic incentives. No preassigned cleavages or factions restrict the coordination between agents, and any subset of the assembly can form a voting bloc. The assembly uses a majority voting rule $r_{\mathcal{N}}$ such that the exogenous policy proposal passes if it gathers at least $r_{\mathcal{N}}N$ votes and an exogenous default policy is implemented otherwise. For notational simplicity, and without loss of generality, I assume that $r_{\mathcal{N}}N$ is an integer.

I assume that the probability distribution over preference profiles Ω has full support, that is, $\Omega(p) > 0$ for all $p \in \{0,1\}^N$. As in the previous section, ex-post utilities are determined by the policy outcome; agent i receives utility one if the outcome is equal to p_i, and zero otherwise.

Agents play a political game with two stages. First, in the coalition formation stage, the agents sort themselves out into voting blocs. Second, in the voting game, agents choose policy by voting.

The coalition formation stage is now more complex. Any subset $C_j \in \mathcal{N}$ may form a voting bloc $V_j = (C_j, r_j)$ with an internal rule characterized by a scaler r_j. Agents choose to join a voting bloc voluntarily, and they may also remain independent. Formally, I assume that there exists a list or finite sequence of available voting rules $(r_0, ..., r_R)$. Rules function as contracts. Agents choose which contract to sign. Let $a_i \in \{0, ..., R\}$ denote the contract or rule chosen by agent i and let $a \equiv (a_1, a_2, ..., a_N)$. For any $j \in \{1, ..., R\}$, the set of agents who choose rule r_j form a voting bloc that aggregates preferences according to rule r_j as described in section 2: For any $C_j = \{i \in \mathcal{N} : a_i = j\}$, the bloc $V_j = (C_j, r_j)$ of size N_j coordinates to vote together with its internal majority if this internal majority given the declared preferences $\hat{p}_{C_j}$ is of size at least $r_j N_j$, and otherwise the bloc casts the same votes in the assembly as those recorded in its internal meeting.

Each agent must sign exactly one contract, but I let $r_0 = 1$ so that agents who choose rule r_0 do not coordinate their actions -technically, I treat them as if they form a voting bloc with unanimity, so they only vote together if they all agree in their preference. I assume that the list of available rules contains enough rules so that for any size N_j and any integer $x \in \left(\frac{N_j}{2}, N_j\right)$, there exists a rule r_j such that $x \leq r_j N_j \leq x + 1$. This implies that for any quota rule, and for any voting blocs of any size, there is a contract available to the bloc that sets the internal rule to be the desired quota rule. Furthermore, I assume that there are enough copies of each voting rule or contract so that several blocs can form, all of them using the same quota rule as their identical internal rule. More precisely, for any k disjoint coalitions $(C_1, ..., C_k)$ that each wishes to form a separate bloc with internal rule $\hat{r}_j$, the sequence of available rules $(r_0, ..., r_R)$ contains a subsequence of length k of the form $(\hat{r}_j, \hat{r}_j, ..., \hat{r}_j)$. Available contracts serve as coordination devices so that agents endogenously choose both their partners and the internal rule of their bloc.

The timing of the game is as follows. First each agent $i \in \mathcal{N}$ simultaneously chooses a contract or rule a_i, while still ignoring her preference p_i and knowing only the common information probability distribution Ω. Second, once agents have partitioned themselves into voting blocs, each $i \in \mathcal{N}$ privately learns p_i, then voting blocs meet and agents vote $\hat{p}_i \in \{0, 1\}$, declaring their preference for or against the proposal. At a third stage, the assembly meets and agents cast votes as determined at the meetings of the different voting blocs. Finally, the policy proposal is implemented if it receives $r_{\mathcal{N}} N$ votes in favor.

I first assume that the vote in the assembly at the third stage is not a strategic choice, since it is mechanically determined by the vote inside each bloc, with full commitment. Under this commitment assumption, whether p and/or $\hat{p}$ become public information after the second stage, or remain undisclosed when the agents meet in the assembly is irrelevant, because there are no choices to be made at the third stage. Only when I relax the commitment assumption in an extension at the end of the section, it becomes crucial that either p or $\hat{p}$ (or both) be revealed before voting takes place in the assembly.

Let $\widehat{p}_i(a, p_i)$ denote the vote of agent i in the internal meeting of her voting bloc, as a function of the partition of the assembly into voting blocs, and the preference of agent i. A strategy s_i for agent i in the whole game is a pair $(a_i, \widehat{p}_i(a, p_i))$, specifying her choice of voting bloc, and her vote in the internal meeting of her bloc. Let $s = (s_1, ..., s_N)$ be a strategy profile for the whole set of agents. Given that all the available rules are $r - majority$ rules or unanimity rule, for any $i \in \mathcal{N}$, any a_i and any s_{-i}, the number of votes cast in the assembly for p_i is weakly greater if $\widehat{p}_i = p_i$ than if $\widehat{p}_i = 1 - p_i$. Thus, if i is ever pivotal, sincere voting at the bloc meeting is a weakly dominant strategy. Note that this claim is stronger than lemma 1, because I treat independents as members of a voting bloc with unanimity, so all the strategic votes occur simultaneously at the second stage. Since deviations from sincerity are dominated, I assume that $\widehat{p}_i(a, p_i) = p_i \; \forall a \in \{0, ..., R\}^N$, $\forall p_i \in \{0, 1\}$, $\forall i \in \mathcal{N}$.

The only remaining strategic consideration is the choice of a voting bloc at the coalition stage. The Nash equilibria of the coalition formation game, together with sincere voting strategies inside the blocs, determine a subgame perfect Nash equilibrium of the whole game. Assuming sincere voting, a pure Nash equilibrium of the voting bloc formation game is a profile of strategies $a^* = (a_1^*, a_2^*, ..., a_N^*)$ such that the choice of which contract to sign by each agent is a best response to the choice by every other agent. It is trivial to find a pure Nash equilibrium: If every agent chooses not to coordinate with anyone else by selecting rule r_0, no agent has an incentive to deviate.

I am interested in pure Nash equilibria in which agents form voting blocs and coordinate their votes.

Definition 3 *A strategy profile s exhibits party discipline if there exists $i \in \mathcal{N}$ such that, with positive probability, $v_i \neq \hat{p}_i$. The profile s exhibits relevant party discipline if either $\left(\sum_{i\in\mathcal{N}} \hat{p}_i < Nr_{\mathcal{N}} \leq \sum_{i\in\mathcal{N}} v_i\right)$ or $\left(\sum_{i\in\mathcal{N}} v_i < Nr_{\mathcal{N}} \leq \sum_{i\in\mathcal{N}} \hat{p}_i\right)$ occurs with positive probability.*

A strategy profile exhibits party discipline if the coordination of votes inside the blocs makes some agent cast a vote against her preference in the assembly. Party discipline is relevant if the outcome in the assembly after the coordination of votes inside the blocs differs from the outcome that would result from no coordination of votes.

There always exist pure Nash equilibria with party discipline, but the existence of pure Nash equilibria with relevant party discipline depends on the rule of the assembly, and the distribution of types of the agents.

Proposition 9 *Suppose Ω has full support and $r_{\mathcal{N}} \in [\frac{N+1}{2N}, \frac{N-1}{N}]$. For any $r_V \leq \frac{N-1}{N}$, there exists a pure strategy Nash equilibrium with party discipline in which $(\mathcal{N}, r_V)$ forms. If $r_V \leq r_{\mathcal{N}} - \frac{1}{N}$, party discipline is relevant.*

Corollary 10 *Suppose Ω has full support and $r_{\mathcal{N}} \in (\frac{N+1}{2N}, \frac{N-1}{N}]$. Then there exists an equilibrium with relevant party discipline.*

The proof is constructive, showing that the formation of a voting bloc by the grand coalition is a Nash equilibrium. If every other member belongs to the voting bloc, no individual member gains anything deviating, because the voting bloc, acting together, is a dictator even after it loses one member. Therefore, outsiders only influence the outcome if the voting bloc does not vote together, in which case the outsider would achieve the same outcome as a member of the bloc. If the grand coalition coordinates the votes of its members with an internal majority rule that sets a higher threshold than the rule of the assembly, it does not affect policy outcomes, merely turning into unanimous votes in the assembly preferences that were lopsided, but not unanimous, in the meeting of the voting bloc. On the other hand, if the coalition chooses an internal rule that is lower than the rule of the assembly, the effect is the same as if the assembly endogenously changes its voting rule, and it affects policy outcomes.

While the constructive proof selects the easiest equilibrium with just one bloc, it suffices to look back at example 2 to find that Nash equilibria with relevant party discipline and several blocs are also possible. To make the example fully consistent with the general model of endogenous party formation in this section, define the available rules $r_1 = r_2 = \frac{5}{9}$. Then, $a = (1, 1, 1, 0, 2, 2, 2)$ together with sincere voting defines a pure Nash equilibrium in which agents $1, 2, 3$ form a voting bloc, agents $5, 6, 7$ form a second voting bloc, and agent 4 remains independent. These voting blocs affect policy outcomes, so the equilibrium exhibits relevant party discipline. The numbers in table 1 show that members of a bloc become worse off if they leave or switch blocs, and that the independent becomes worse off if she joins a bloc.

If $r_{\mathcal{N}}$ is simple majority, pure Nash equilibria with relevant party discipline are not guaranteed to exist. Here is a counterexample.

Example 4 *Let $N = 7$, $r_{\mathcal{N}} = \frac{4}{7}$ and suppose that types are independent and $t_i = 0.5$ for any $i \in \mathcal{N}$. Then there is no pure Nash equilibrium with relevant party discipline. If a unique bloc of size three forms, outsiders prefer to join it (utility $9/16$ outside the bloc, $11/16$ joining); if two blocs of size three form, each member of a bloc is better off switching blocs to create a unique bloc of size four ($11/16$ compared to $10/16$ with two blocs); and if a bloc of size four or more forms, this bloc acts as a dictator, so every outsider is better off joining it. It follows that in equilibrium, either no blocs form, or the grand coalition forms a unique voting bloc, which does not affect the policy outcome.*

If $r_{\mathcal{N}}$ is simple majority and there is a pure Nash equilibrium with relevant party discipline, the size of the blocs in this equilibrium must be less than minimal winning.

Proposition 11 *Let $r_{\mathcal{N}} = \frac{N+1}{2N}$. Then in any pure Nash equilibrium with relevant party discipline, $N_j < \frac{N+1}{2}$ for any voting bloc (C_j, r_j) with a simple majority internal voting rule, and if there exists at least one agent who remains independent, then $N_k < \frac{N+1}{2}$ for any voting bloc (C_k, r_k) with any majority rule $r_k \leq \frac{N-1}{2N_K}$.*

A voting bloc smaller than the grand coalition cannot be large enough to act as a dictator in equilibrium. If it is large enough to act as a dictator, every agent would like to join until the grand coalition forms an irrelevant voting bloc. With no barriers to enter blocs, competition among opposing blocs only occurs if the weaker blocs also have a hope of influencing the policy outcome. If no single bloc is large enough to act as a dictator, each voting bloc affects the outcome with positive probability, and agents have incentives to coordinate in voting blocs.

A complete characterization of the number of blocs that may form in equilibrium for any arbitrary assembly is not tractable, but under some restrictions on the structure of preferences, I show that if the assembly is sufficiently polarized, then a pure Nash equilibrium with relevant party discipline and two voting blocs with simple majority internal voting rules exists.

Condition 1 *Let* $K = \frac{N-1}{2}$. *Let* $\gamma_1, ..., \gamma_K$ *be strictly increasing in* k *and such that* $\gamma_1 > 0$, *and let* $\lambda \in \left(0, \frac{1}{2\gamma_K}\right)$. *Suppose types are independent and symmetrically distributed, so that* n_0 *agents have type* $\frac{1}{2}$, *and for each* $k \in \{1, ...K\}$, n_k *agents have type* $t_{k-} = 0.5 - \lambda\gamma_k$ *and* n_k *agents have type* $t_{k+} = 0.5 + \lambda\gamma_k$, *with* $n_0 + \sum_{k=1}^{K} 2n_k = N$.

This condition is satisfied by any symmetric and independent distribution of preferences, and it is notationally convenient to describe a measure of the polarization of the preferences in the assembly by the parameter λ; increasing λ stretches out the distribution of probabilities of favoring the proposal, reducing the individual uncertainty of the agents. If the distribution of preferences satisfies this condition, there is an equilibrium in which the assembly partitions into two voting blocs and a subset of independent agents. This is not a knife-edge result, rather, it holds for types that are not independent nor symmetric, as long as they are sufficiently close to satisfying Condition 1.

Let $\mathbf{\Omega}$ be the set of all possible probability distributions over preference profiles with full support. Let Ω, Ω' be arbitrary elements of $\mathbf{\Omega}$**.** Let $d(\Omega, \Omega') = \max_{p \in \{0,1\}^N} \Omega(p) - \Omega'(p)$ be a metric so that $(\mathbf{\Omega}, d)$ is a metric space. Let $O_\lambda(\Omega)$ be the open $\lambda-$neighborhood around Ω in this space.

Proposition 12 *Suppose* Ω *satisfies condition 1,* $n_K \geq 3$ *and* $r_{\mathcal{N}} = \frac{N+1}{2N}$. *There exist* $\lambda^* \in \left(0, \frac{1}{2\gamma_K}\right)$ *and* $\Omega \in \mathbf{\Omega}$ *such that for any* $\lambda > \lambda^*$ *and any* $\Omega' \in O_\lambda(\Omega)$, *a pure Nash equilibrium with relevant party discipline and two simple-majority voting blocs exists.*

Example 2 discussed above illustrates the intuition for this result. If the assembly is sufficiently polarized, and the types of the agents are independent and symmetric or almost symmetric, the subset of agents at the extremes of the distribution of types can each form a voting bloc, such that the low-type bloc is so likely to vote against the proposal that agents with a higher type do not want to join it, the high-type bloc is so likely to vote for the proposal that agents with a lower type do not want to join it, and members of a bloc do not want to switch over to the other bloc.

Equilibria with multiple blocs are also possible, at least for some configurations of types. In fact, the supremum on ratio of blocs to size of the society over all N and all Ω is $\frac{1}{3}$. This follows readily as a corollary from the following generalized example, which I state and proof as a claim.

Claim 13 *Given any arbitrary* $K \in \mathbb{N}$, *assume* $N = 6K + 1$. *For any* $\varepsilon > 0$, *let* Ω_ε *be such that agents have independent types,* $t_i = \varepsilon \ \forall i \in \{1, ..., 3K\}$, $t_i = 1 - \varepsilon \ \forall i \in \{3K + 1, ..., 6K\}$ *and* $t_N = \frac{1}{2}$. *There exist* $\varepsilon^* > 0$ *and* $\lambda_\varepsilon : (0, \varepsilon^*) \longrightarrow (0, \varepsilon^*)$ *such that for any* $\varepsilon \in (0, \varepsilon^*)$, *for any* $\lambda \in (0, \lambda_\varepsilon(\varepsilon))$, *and for any* $\Omega' \in O_\lambda(\Omega_\varepsilon)$, *there is an equilibrium with relevant party discipline with* $2K$ *voting blocs. In this equilibrium, for each* $k \in \{1, ..., 2K\}$, *agents* $\{3k - 2, 3k - 1, 3k\} \subseteq \mathcal{N}$ *form a voting bloc with simple majority as internal voting rule.*

While claim 13 suggests that the number of voting blocs can be very large, note that if many blocs form, these blocs face incentives to further merge with other blocs to form coalitions of blocs. We could then apply the theory taking blocs to be the agents who engage in coalition formation in a second level of structure, so that individual agents form factions (blocs), and factions form alliances (blocs of blocs). I leave this extension to future research.

The set of admissible probability distributions over preferences in the theory is too rich to characterize the set of distributions for which a specific class of equilibria exists. The equilibria of the game are highly sensitive to the specific probability distribution of preference profiles.[8] From an applied perspective, a more promising avenue is to gather information on the actual probability distribution over preferences on any specific assembly of interest, and then use the theory to predict the formation of parties within this assembly.

3.1 Robustness

An open theoretical question is the extent to which the theory is robust to alternative protocols of party formation, and to the relaxation of the assumption of commitment. I consider other solution concepts and games of coalition formation first, and the case with no commitment at the end of the subsection.

In section 3, a member of a party can deviate to join another party; the doors of parties are fully open to outsiders. On the other hand, in section 2, the doors of a party are closed to outsiders. A general class of solution concepts that contains these two as special cases, allows for entry into a party if a fraction $\alpha \in \mathbb{R}_+$ of the members of the party benefit from the entry of the new member. The case with $\alpha > 1$ means that entry is impossible; whereas $\alpha = 0$ corresponds to open entry. While any value of $\alpha \in (0, 1)$ may have some justification, perhaps the third most natural alternative is $\alpha = 1$, so that a member may enter by unanimous consent. This

[8] For instance, it is possible to construct an example (available from the author) in which there is an equilibrium with relevant party discipline with two parties, but there is no equilibrium with relevant party discipline with a single party.

is the admission rule into the European Union, and it correspond to the notion of *individual stability* due to Drèze and Greenberg (1980).

To model voting blocs that restrict entry, I use cooperative game theory, with the partition function approach. Let a voting bloc structure (π, r) be a pair consisting of a partition of the assembly $\pi = (\pi_0, ..., \pi_k)$ and a list of voting rules $r = (r_0, ..., r_k)$, one voting rule associated to each element of the partition. An element π_j, together with its rule r_j, is a voting bloc. A stable voting bloc is one that no agent wishes to leave for any other bloc where the agent would be accepted. Assume agents vote sincerely. Then the payoffs are solely determined by the voting bloc structure.

Definition 4 *A voting bloc structure (π, r) is $\alpha-$stable if for any $\pi_j \in \pi$ and for any $i \in \mathcal{N}$ such that $i \in \pi_j$,*

(i) Agent i is weakly worse off if she leaves π_j to become an independent.

(ii) Agent i is weakly worse off if she leaves π_j to join π_h for any $h \in \{1, ..., k\}$ such that the fraction of π_h-members who become weakly better off after i joins is at least α.

The solution concept of $0-$stability corresponds to the Nash equilibrium of the game in section 3. There is an ordering among all the solution concepts in the class of $\alpha-$stability. Since increasing alpha allows fewer deviations, it follows that for any $\alpha' > \alpha$, the set of $\alpha-$stable voting bloc structures is contained in the set of $\alpha'-$stable voting bloc structures. A corollary of this observation, together with proposition 9, is that for any α, if Ω has full support, there exists an $\alpha-$stable voting bloc structure with party discipline, and if the rule in the assembly is not simple majority, then party discipline is relevant. For $\alpha \geq 1$, the result can be strengthened to encompass simple majority.

Proposition 14 *Suppose Ω has full support and $r_{\mathcal{N}} = \frac{N+1}{2N}$. For $\alpha \geq 1$, there exists an $\alpha-$stable voting bloc structure with relevant party discipline.*

The proof is constructive, letting a coalition of $N - 2$ agents form a voting bloc with simple majority. This voting bloc is a dictator even if it loses one member, so no member gains anything deviating. The two outsiders would prefer to enter, but if they enter, they create a net loss in utility for the bloc, which means that at least one member is hurt by their entrance, so they are vetoed.

For any $\alpha > 1$, $\alpha-$stability corresponds to the Nash equilibrium of a well known game in the non-cooperative coalition formation literature: the exclusive membership game named Δ by Hart and Kurz (1983), in which each agent announces a coalition that she would like to join, and all the agents who make the same announcement form a coalition, even if the resulting coalition is just a strict subset of the announced one. This game implies that if an agent leaves a bloc, the remaining members of the bloc stay together. In an alternative exclusive membership game, labeled Γ by Hart and Kurz (1983), each agent announces a coalition, and each announced coalition forms only if all its members make the same announcement. This is the game Γ in the notation of Hart and Kurz (1983). It implies that if a member leaves, it triggers the dissolution of the bloc.

The game in section 3 is an open membership coalition formation game in which every agent can deviate to join another coalition. Some of the results are robust if agents play an exclusive membership game instead. Since every equilibrium of the game in section 3 is an equilibrium of the game Δ, it follows as a corollary that propositions 9, 12 and claim 13 on the existence of equilibria with one or more parties, with two parties, and with many parties hold for game Δ. For game Γ, the existence of equilibria with party discipline in proposition 9 holds, but the proof of existence of relevant party discipline does not hold, because some members may want to dissolve the bloc if dissolving it can change the policy outcome. Proposition 12 and claim 13 on the existence of equilibria with two parties and with many parties hold as well in the game Γ, because every bloc members benefit (approximately) equally from her bloc, so there is no incentive to dissolve it. On the other hand, proposition 11 on the size of parties does not hold in any exclusive membership game. A minimal winning bloc in an assembly with identical agents can be sustained as an equilibrium. The intuition for proposition 11 and the non existence of large blocs is that if a bloc is a dictator, then every other agent would deviate to join it. These deviations are not feasible in an exclusive membership game.

We may also consider a more complex institutional set up, in which agents move sequentially. Suppose blocs do not form spontaneously, but rather, there exists a subset of agents who are leaders. Leaders have the ability to send out invitations to any other agents, and those who accept the invitation of a leader, form a voting bloc. This set up is also an exclusive membership game that leads to similar results as the exclusive membership games Δ and Γ, but with a caveat that may be important in applications: The formation of a bloc now depends on the existence of a leader within the membership of that bloc. For instance, proposition 12 holds only as long as one of the members of the two blocs that form is a leader who can coordinate her fellow bloc members. If leaders hold a monopoly on the technology of coordination, in applications we may observe assemblies with distributions of preferences that are ripe for the formation of blocs, but in which blocs fail to materialize due to the absence of a leader.

The solution concepts and games considered so far only allow individual deviations. However, it is plausible that agents not only can coordinate to form a voting bloc, but they can also coordinate a deviation to leave a bloc or form a new one. Further borrowing from cooperative notions, the most robust solution concept is the core, such that given a voting bloc structure, no subset of agents have an incentive to deviate to form a new voting bloc. Several definitions of the core exist, depending on the expected reaction by the other agents after a deviation (Hart & Kurz 1983). Under any of these definitions, the core is easily empty. Consider a rather simple version of the core, in which other agents do not react to the deviation, continuing to coordinate in their respective voting blocs. Under this definition, both the small polarized assembly in example 2 and the simplistic example 4 have an empty core.[9]

Leaving for future research further extensions of the theory to other coalition formation games and solution concepts, I return to the benchmark open membership endogenous coalition formation game in section 3, now assuming that there is no commitment. Blocs must then deal with the ex-post incentives to defect. As in the extension in subsection 2.1, assume that the internal meeting of a voting bloc does not determine the votes of the bloc in the assembly; it merely issues a recommendation. Voting in the assembly is a strategic choice and members choose whether to follow the recommendation of their bloc or to ignore it, in which case they incur a small cost $c \geq 0$.

Every bloc can anticipate that coordination will fail if it leads to a marginal outcome decided by a single vote. With this difficulty in mind, and assuming that either the true preference profile p or the declared preferences profile $\hat{p}$ becomes public before the vote in the assembly, blocs can use recommendation rules that only require vote members to coordinate if the coordination by all blocs leads to an outcome decided by more than one vote. This voting recommendation is not enforceable, but since no agent is individually pivotal, agents do not have an incentive to deviate from the voting prescription in their bloc, and thus an arbitrarily small cost $c > 0$ suffices to make agents follow the recommendations. Therefore voting blocs can coordinate their votes and change the policy outcome in any event in which the coordination leads to an outcome decided by more than one vote. Without commitment, parties only lose the power to enforce party discipline in the events when an individual agent is pivotal. The proof of proposition 9 on the existence of Nash equilibria with relevant party discipline constructs an equilibrium in which the outcome is never decided by one vote; therefore, it holds as stated in a game without enforcement.

[9] Not all deviations allowed in the core definition of stability are equally plausible. In a related paper (Eguia forthcoming), I apply the model to a stylized assembly with nine agents, and in this generalized example I consider of a solution concept that allows agents to coordinate deviations with other members of their bloc, but not across blocs.

4 Conclusion

I have shown that voting blocs form in equilibrium in a model in which agents with heterogeneous preferences coalesce into voting blocs.

In a model with two parties that can each form a voting bloc, I have shown the necessary and sufficient condition for every member in a party to have an incentive to join the bloc, and how these incentives change with variations on the type of the agents, the voting rule chosen by the parties, the sizes of the parties and the polarization of the assembly.

In a model with no pre-existing parties, I have shown the coordination of votes for an ideological gain is sufficient motivation for agents to coalesce into voting blocs, coordinating their actions to vote together and affecting the policy outcome. I have shown that if the preferences in the assembly are sufficiently polarized, there is an equilibrium in which two voting blocs form, one at each side of the ideological divide, both of size less than minimal winning, and I have sketched how the results generalize to the case with no commitment, and to variations in the protocol of the coalition formation game.

5 Appendix

Proof of Lemma 1

Proof. Suppose i does not join a voting bloc. Let $k = \#\{j \in \mathcal{N}\backslash\{i\} : v_j = p_i\}$. If $k \neq \frac{N-1}{2}$, then the policy outcome is the same regardless of the vote by i. If $k = \frac{N-1}{2}$, then the outcome coincides with the preference of i if and only if $v_i = p_i$, and thus in this case i is strictly better off voting sincerely. Therefore, sincere voting is weakly dominant for i.

Suppose i joins a voting bloc (C, r_J) with size N_C. Let $k = \#\{j \in C\backslash\{i\} : \widehat{p}_j = p_i\}$. If agents who do not join a voting bloc follow their weakly dominant strategy and vote sincerely, the probability that the policy outcome coincides with the preference of i is non decreasing in the number of votes cast in the assembly by bloc C for the outcome preferred by i. If $k \notin [N_C - \lceil r_C N_C \rceil, \lceil r_C N_C \rceil - 1]$, then the policy outcome is the same regardless of the vote by i inside the internal meeting of C. If $k = N_C - \lceil r_C N_C \rceil$, then $\#\{j \in C : v_j = p_i\} = 0$ if $\widehat{p}_i \neq p_i$ and $\#\{j \in C : v_j = p_i\} = N_C - \lceil r_C N_C \rceil + 1$ if $\widehat{p}_i = p_i$, thus i is (weakly) better off in expectation by voting $\widehat{p}_i = p_i$. If $k \in [N_C - \lceil r_C N_C \rceil + 1, \lceil r_C N_C \rceil - 2]$, then $\#\{j \in C : v_j = p_i\} = k$ if $\widehat{p}_i \neq p_i$ and $\#\{j \in C : v_j = p_i\} = k + 1$ if $\widehat{p}_i = p_i$, thus i is (weakly) better off in expectation by voting $\widehat{p}_i = p_i$. If $k = \lceil r_C N_C \rceil - 1$, then $\#\{j \in C : v_j = p_i\} = \lceil r_C N_C \rceil - 1$ if $\widehat{p}_i \neq p_i$ and $\#\{j \in C : v_j = p_i\} = N_C$ if $\widehat{p}_i = p_i$, thus i is (weakly) better off in expectation by voting

$\widehat{p}_i = p_i$. Therefore, in every event i is either indifferent or strictly better off voting sincerely. ■

Proof of Proposition 2

To prove this proposition I first prove two technical claims, and a lemma. For all results in the proof of proposition 2, assume that types are independent. The following notation becomes useful: Let $g^C(x) \equiv P\left[\sum_{i \in C} p_i = x\right]$.

First I show that if a coalition C of size N_C leans right, then the distribution of the number of agents in C who favor the proposal is such that for any size of the majority inside C, it is more likely that the majority is in favor than against the proposal. Second, I show that if M is symmetric and L leans left and forms a voting bloc, then the distribution of the number of votes cast by $L \sqcup M$ in the division of the assembly is such that given any absolute difference between the number of votes $L \sqcup M$ casts for and against the proposal, the net difference is negative with probability at least a half. Readers who wish to skip the proof of these technical claims can read ahead to Lemma 17.

Let $\lfloor x \rfloor \equiv \max\{k \in \mathbb{Z} : k \leq x\}$ be the floor function, which gives the largest integer less than or equal to x.

Claim 15 *Assume C leans right. Then*

$$g^C(k) \leq g^C(N_C - k) \ \forall k \in \left\{0, 1, ..., \left\lfloor \frac{N_C}{2} \right\rfloor \right\}. \tag{1}$$

Proof. Label every $i \in C$ from 1 to N_C by type so that $t_i \leq t_j$ implies $i \leq j$. If N_C is odd, let m denote agent $\frac{N_C+1}{2}$ with the median type in coalition C. Let $D \subseteq C$ be equal to $C \backslash \{m\}$ if N_C is odd, and equal to C if C is even. I first show that

$$g^D(k) \leq g^D(N_D - k) \tag{2}$$

for any $k \in \left\{0, 1, ..., \frac{N_D}{2}\right\}$.

For each k, $g^D(k)$ is equal to the summation of $\binom{N_D}{k}$ terms. Each of these terms is the probability of the event in which a particular set of k agents in D favor the proposal, and every other agent in D opposes it. I index these events. For any $\tau \in \left\{1, ..., \binom{N_D}{k}\right\}$, let D_τ be the set of k agents in favor of the proposal in the $\tau - th$ event and let $D'_\tau = \{i \in D : i \notin D_\tau$ and $N_C + 1 - i \in D_\tau\}$. Let ρ_τ be the probability of the $\tau - th$ event. Note that $|D_\tau| - |D'_\tau|$ is even and that

$$\rho_\tau = \prod_{i \in D_\tau} t_i \prod_{j \in D \backslash D_\tau} (1 - t_j) = \prod_{i \in D_\tau} t_i \prod_{j \in D'_\tau} (1 - t_j) \prod_{j \in D \backslash (D_\tau \sqcup D'_\tau)} (1 - t_j).$$

Similarly, $g^D(N_D - k)$ is equal to the summation of $\binom{N_D}{k}$ terms, where each term is the probability of the event in which a given set of $N_D - k$ favor the proposal and all other agents in D oppose it. Construct an index of these events from $\binom{N_D}{k}+1$ to $2\binom{N_D}{k}$ based on the index of the events from 1 to $\binom{N_D}{k}$, as follows. Let $\nu = \tau - \binom{N_D}{k}$. For any $\tau \in \left\{\binom{N_D}{k}+1, ..., 2\binom{N_D}{k}\right\}$, event τ is such that in this event the k agents in D_ν favor the proposal, all the agents in D'_ν oppose it, and for any $i \notin D'_\nu$ who opposes the proposal in event τ, agent $N_C + 1 - i$ opposes it as well. This index of events from $\binom{N_D}{k}+1$ to $2\binom{N_D}{k}$ is not unique; arbitrarily choose one index that satisfies the above condition. For any $\tau \in \left\{\binom{N_D}{k}+1, ..., 2\binom{N_D}{k}\right\}$, let D_τ be the set of $N_D - k$ agents in favor of the proposal, let D'_τ be defined as before (note $D'_\tau = D'_\nu$ by construction), and let D''_τ be the set of agents who oppose the proposal in event τ but are not members of D'_τ. Then for any $\tau \in \left\{\binom{N_D}{k}+1, ..., 2\binom{N_D}{k}\right\}$

$$\rho_\tau = \prod_{i\in D_\nu} t_i \prod_{j\in D'_\nu}(1-t_j)\prod_{j\in D''_\tau}(1-t_j)\prod_{i\in D\backslash(D_\nu\sqcup D'_\nu\sqcup D''_\tau)} t_i.$$

To prove inequality 2, it suffices to show that for any $\tau \in \left\{\binom{N_C-1}{k}+1, ..., 2\binom{N_C-1}{k}\right\}$,

$$\begin{aligned}
\rho_\nu &\leq \rho_\tau \\
\prod_{j\in D\backslash(D_\nu\sqcup D'_\nu\sqcup D''_\tau)}(1-t_j) &\leq \prod_{i\in D\backslash(D_\nu\sqcup D'_\nu\sqcup D''_\tau)} t_i \\
\prod_{j\in D\backslash(D_\nu\sqcup D'_\nu\sqcup D''_\tau)}(1-t_j) &\leq \prod_{j\in D\backslash(D_\nu\sqcup D'_\nu\sqcup D''_\tau)} t_{N_C+1-j},
\end{aligned}$$

where the first inequality implies the second by cancelling out terms that are common to each side of the inequality, and the second inequality implies the third because for any $\tau \in \left\{\binom{N_D}{k}+1, ..., 2\binom{N_D}{k}\right\}$, the sets $(D_\nu \sqcup D'_\nu)$ and D''_τ are sets of agents in symmetric positions around the median in the order of agents in C according to types, and thus $D\backslash(D_\nu \sqcup D'_\nu \sqcup D''_\tau)$ is also a set of agents in symmetric positions. Finally, the third inequality holds because for each j,

$$1 - t_j \leq t_{N_C+1-j}$$

by assumption.

It remains to be shown that if N_C is odd, once we add the median m to D, for any $k \in \{0, 1, ..., \frac{N_C-1}{2}\}$

$$g^C(k) \leq g^C(N_C - k)$$

holds as well. Note that since C leans right, $t_m \geq \frac{1}{2}$.

More generally than necessary for this proof, I next show that for any $i \notin C$ with $t_i \geq \frac{1}{2}$, any C and any $k \in \left\{0, 1, ..., \left\lfloor \frac{N_C}{2} \right\rfloor\right\}$, if inequality 1 holds for C, then inequality 1 holds for

$C \sqcup \{i\}$. I want to show

$$g^{C\sqcup\{i\}}(k) \leq g^{C\sqcup\{i\}}(N_C + 1 - k) \quad (3)$$
$$t_i g^C(k-1) + (1-t_i) g^C(k) \leq t_i g^C(N_C - k) + (1-t_i) g^C(N_C + 1 - k). \quad (4)$$

As shown by Darroch (1964), the distribution of successes in N Bernouilli trials is unimodal. Thus, if $g^C(N_C + 1 - k) \geq g^C(N_C - k)$, the proof is complete. If instead $g^C(N_C + 1 - k) < g^C(N_C - k)$, by unimodality, note that inequality 1 implies

$$g^C(k) \geq g^C(k-1).$$

Then, $t_i \geq \frac{1}{2}$ implies that the left hand side of inequality 4 is less than or equal to

$$(1-t_i) g^C(k-1) + t_i g^C(k)$$

and thus to prove inequality 4, it suffices to show that

$$(1-t_i) g^C(k-1) + t_i g^C(k) \leq t_i g^C(N_C - k) + (1-t) g^C(N_C + 1 - k),$$

which is true by inequality 1. ■

Claim 16 *Let $\mathcal{N} = L \sqcup M \sqcup R$. Assume M is symmetric, and L leans left and forms a voting bloc (L, r_L). Then:*

$$P\left[\sum_{i \in L\sqcup M} v_i = \frac{N_L + N_M}{2} - k\right] \geq P\left[\sum_{i \in L\sqcup M} v_i = \frac{N_L + N_M}{2} + k\right] \text{ for any positive } k.$$

Proof. As a preliminary step, note that by claim 15, since M is symmetric and L leans left, $g^M(\frac{N_M - 1}{2} - k) = g^M(\frac{N_M + 1}{2} + k)$ and $g^L(\frac{N_L - 1}{2} - k) \geq g^L(\frac{N_L + 1}{2} + k)$ for any non negative integer k.

Define L to be *active* given a preference profile p if it reverses the votes of its internal minority given the rule r_L, so that $v_i \neq p_i$ for some $i \in L$ and let us define L to be *inactive* otherwise. Then the probability that $L \sqcup M$ casts x votes in favor of the policy proposal is

$$P\left[\sum_{i \in L\sqcup M} v_i = x | L \text{ } active\right] P[L \text{ } active] + P\left[\sum_{i \in L\sqcup M} v_i = x | L \text{ } inactive\right] P[L \text{ } inactive].$$

I first want to show that

$$P\left[\sum_{i \in L\sqcup M} v_i = \frac{N_L + N_M}{2} - k | L \text{ } active\right] \geq P\left[\sum_{i \in L\sqcup M} v_i = \frac{N_L + N_M}{2} + k | L \text{ } active\right]. \quad (5)$$

Noting that if L is active then $\sum_{i \in L} v_i \in \{0, N_L\}$, that $\sum_{i \in M} v_i = \sum_{i \in M} p_i$ for all preference profiles, and that

$$g^M \left(\frac{N_M - 1}{2} - k \right) = g^M \left(\frac{N_M + 1}{2} + k \right)$$

for any k, rewrite inequality 5 as:

$$P[\sum_{i \in L} v_i = N_L | L \ active] g^M \left(\frac{N_M - N_L}{2} - k \right) + P[\sum_{i \in L} v_i = 0 | L \ active] g^M \left(\frac{N_M + N_L}{2} - k] \right)$$
$$\geq P[\sum_{i \in L} v_i = N_L | L \ active] g^M \left(\frac{N_M + N_L}{2} - k \right) + P[\sum_{i \in L} v_i = 0 | L \ active] g^M \left(\frac{N_M - N_L}{2} - k \right).$$

Regrouping terms:

$$\left(P[\sum_{i \in L} v_i = N_L | L \ active] - P[\sum_{i \in L} v_i = 0 | L \ active] \right)$$
$$\left(g^M \left(\frac{N_M - N_L}{2} - k \right) - g^M \left(\frac{N_M + N_L}{2} - k] \right) \right) \geq 0.$$

Since L leans left, the first term is weakly negative; since the distribution of the number of agents in M who favor the policy proposal is symmetric (and unimodal), the second term is also negative. Thus the expression is weakly positive, as desired.

Second, I want to show that

$$P\left[\sum_{i \in L \sqcup M} v_i = \frac{N_L + N_M}{2} - k | L \ inactive \right] \geq P\left[\sum_{i \in L \sqcup M} v_i = \frac{N_L + N_M}{2} + k | L \ inactive \right]. \quad (6)$$

Note that L is inactive if and only if $(1 - r_L) N_L < \sum_{i \in L} p_i < r_L N_L$.

$$P\left[\sum_{i \in L \sqcup M} v_i = \frac{N_L + N_M}{2} + k | L \ ina. \right] - P\left[\sum_{i \in L \sqcup M} v_i = \frac{N_L + N_M}{2} - k | L \ ina. \right] =$$
$$\sum_{h=0}^{r_L N_L - \frac{N_L+3}{2}} \left\{ \begin{array}{c} P[\sum_{i \in L} p_i = \frac{N_L+1}{2} + h | L \ ina.] \left(g^M(\frac{N_M-1}{2} + k - h) - g^M(\frac{N_M-1}{2} - k - h) \right) + \\ P[\sum_{i \in L} p_i = \frac{N_L-1}{2} - h | L \ ina.] \left(g^M(\frac{N_M+1}{2} + k + h) - g^M(\frac{N_M+1}{2} - k + h) \right) \end{array} \right\}$$
$$= \sum_{h=0}^{r_L N_L - \frac{N_L+3}{2}} \left\{ \begin{array}{c} \left(P[\sum_{i \in L} p_i = \frac{N_L+1}{2} + h | L \ inactive] - P[\sum_{i \in L} p_i = \frac{N_L-1}{2} - h | L \ inactive] \right) \\ \left(g^M(\frac{N_M-1}{2} + k - h) - g^M(\frac{N_M-1}{2} - k - h) \right) \end{array} \right\}.$$

For any h and k, the first parenthesis is negative because L leans left, and the second one is positive because the distribution of the number of agents in M who favor the policy proposal is unimodal and symmetric around $N_M/2$. Thus the whole expression is negative and inequality (6) holds as desired. ■

I now state and prove lemma 17. For any set C and any two agents $i, j \in C$, let $C_{-ij} \equiv C \backslash \{i, j\}$.

Lemma 17 *Let $\mathcal{N} = L \sqcup M \sqcup R$. Assume $\lceil r_J(N_J - 1)\rceil = \lceil r_J N_J \rceil$ for $J \in \{L, R\}$, types are independent, L leans left and forms a voting bloc (L, r_L), M is symmetric and R leans right. If $a_l = 1$ is a best response for l_R when every other member of party R joins the bloc with rule r_R, then $a_j = 1$ $\forall j \in R$ is a mutual best response $\forall j \in R$.*

Proof. For clarity within this proof, let l unambiguously denote l_R. Since R leans right, R_{-lh_R} leans right as well, and thus R_{-lj} leans right for any $j \in R_{-l}$. Given that every other member but j joins the bloc, member $j \in R$ prefers to participate in the voting bloc (R, r_R) if and only if $t_j(\pi_j^{1+} - \pi_j^{1-}) + (1 - t_j)(\pi_j^{0+} - \pi_j^{0-}) > 0$. Suppose $t_l(\pi_l^{1+} - \pi_l^{1-}) + (1 - t_l)(\pi_l^{0+} - \pi_l^{0-}) \geq 0$. We want to show that

$$t_j(\pi_j^{1+} - \pi_j^{1-}) + (1 - t_j)(\pi_j^{0+} - \pi_j^{0-}) - t_l(\pi_l^{1+} - \pi_l^{1-}) - (1 - t_l)(\pi_l^{0+} - \pi_l^{0-}) \geq 0,$$

which implies $t_j(\pi_j^{1+} - \pi_j^{1-}) + (1 - t_j)(\pi_j^{0+} - \pi_j^{0-}) \geq 0$.

Let

$$P_1 \equiv P\left[\sum_{i \in L \sqcup M} v_i \in [\frac{N_M + N_L}{2} - \frac{N_R - 1}{2}, \frac{N_M + N_L}{2} + \frac{N_R - 1}{2} - \lceil r_R N_R \rceil]\right],$$

$$P_2 \equiv P\left[\sum_{i \in L \sqcup M} v_i \in [\frac{N_M + N_L}{2} - \frac{N_R - 1}{2} + \lceil r_R N_R \rceil, \frac{N_M + N_L}{2} + \frac{N_R - 1}{2}]\right],$$

$$\begin{aligned} P_3 &\equiv P\left[\sum_{i \in L \sqcup M} v_i = \frac{N_L + N_M}{2} + \frac{N_R - 1}{2}\right], \\ \text{and } P_4 &\equiv P\left[\sum_{i \in L \sqcup M} v_i = \frac{N_M + N_L}{2} - \frac{N_R - 1}{2}\right]. \end{aligned}$$

Then, $t_j(\pi_j^{1+} - \pi_j^{1-}) + (1 - t_j)(\pi_j^{0+} - \pi_j^{0-})$ is equal to:

$$\begin{aligned} & t_j \left[\begin{array}{c} \left(t_l P[\sum_{i \in R_{-lj}} p_i = \lceil r_R N_R \rceil - 2] + (1 - t_l) P[\sum_{i \in R_{-lj}} p_i = \lceil r_R N_R \rceil - 1] \right) P_1 \\ - \left(t_l P[\sum_{i \in R_{-lj}} p_i \leq \lfloor (1 - r_R) N_R \rfloor - 2] + (1 - t_l) P[\sum_{i \in R_{-lj}} p_i \leq \lfloor (1 - r_R) N_R \rfloor - 1] \right) P_3 \end{array} \right] \\ & + (1 - t_j) \left[\begin{array}{c} \left(t_l P[\sum_{i \in R_{-lj}} p_i = \lfloor (1 - r_R) N_R \rfloor - 1] + (1 - t_l) P[\sum_{i \in R_{-lj}} p_i = \lfloor (1 - r_R) N_R \rfloor] \right) P_2 \\ - \left(t_l P[\sum_{i \in R_{-lj}} p_i \geq \lceil r_R N_R \rceil - 1] + (1 - t_l) P[\sum_{i \in R_{-lj}} p_i \geq \lceil r_R N_R \rceil] \right) P_4 \end{array} \right], \end{aligned} \tag{7}$$

and $t_l(A_l^+ - A_l^-) + (1-t_l)(B_l^+ - B_l^-)$ is equal to

$$t_l \left[\begin{array}{c} \left(t_j P[\sum_{i\in R_{-lj}} p_i = \lceil r_R N_R \rceil - 2] + (1-t_j) P[\sum_{i\in R_{-lj}} p_i = \lceil r_R N_R \rceil - 1] \right) P_1 \\ - \left(t_j P[\sum_{i\in R_{-lj}} p_i \leq \lfloor (1-r_R) N_R \rfloor - 2] + (1-t_j) P[\sum_{i\in R_{-lj}} p_i \leq \lfloor (1-r_R) N_R \rfloor - 1] \right) P_3 \end{array} \right]$$

$$+ (1-t_l) \left[\begin{array}{c} \left(t_j P[\sum_{i\in R_{-lj}} p_i = \lfloor (1-r_R) N_R \rfloor - 1] + (1-t_j) P[\sum_{i\in R_{-lj}} p_i = \lfloor (1-r_R) N_R \rfloor] \right) P_2 \\ - \left(t_j P[\sum_{i\in R_{-lj}} p_i \geq \lceil r_R N_R \rceil - 1] + (1-t_j) P[\sum_{i\in R_{-lj}} p_i \geq \lceil r_R N_R \rceil] \right) P_4 \end{array} \right]. \tag{8}$$

Therefore $t_j(A_j^+ - A_j^-) + (1-t_j)(B_j^+ - B_j^-) - t_l(A_l^+ - A_l^-) - (1-t_l)(B_l^+ - B_l^-)$ is equal to

$$(t_j - t_l) \left(\begin{array}{c} P[\sum_{i\in R_{-lj}} p_i = \lceil r_R N_R \rceil - 1] P_1 - P[\sum_{i\in R_{-lj}} p_i = \lfloor (1-r_R) N_R \rfloor - 1] P_2 \\ + P[\sum_{i\in R_{-lj}} p_i \geq \lceil r_R N_R \rceil - 1] P_4 - P[\sum_{i\in R_{-lj}} p_i \leq \lfloor (1-r_R) N_R \rfloor - 1] P_3 \end{array} \right). \tag{9}$$

Since M is symmetric and L leans left, it follows by claims 15 and 16 that $P_1 \geq P_2$ and $P_3 \leq P_4$. Since R_{-lj} leans right,

$$\begin{aligned} P[\sum_{i\in R_{-lj}} p_i &= \lceil r_R N_R \rceil - 1] \geq P[\sum_{i\in R_{-lj}} p_i = \lfloor (1-r_R) N_R \rfloor - 1] \text{ and} \\ P[\sum_{i\in R_{-lj}} p_i &\geq \lceil r_R N_R \rceil - 1] \geq P[\sum_{i\in R_{-lj}} p_i \leq \lfloor (1-r_R) N_R \rfloor - 1] \end{aligned}$$

by claim 15. Therefore expression (9) above is non negative as desired. ■

Using lemma 17, I now prove proposition 2.

Proof. By lemma 17, if l_R prefers to participate in the voting bloc, every member of R does. Assuming that the bloc (L, r_L) forms in equilibrium, the bloc (R, r_R) forms as well if and only if l_R wants to participate in the bloc. By analogous reasoning, assuming that (R, r_R) forms in equilibrium, the bloc (L, r_L) forms as well if and only if h_L want to participate in the bloc. Therefore, there exists an equilibrium in which both blocs form if and only if both h_L and l_R want to join in. I show the equilibrium condition for l_R to join. The condition for h_L is analogously derived. Let l unambiguously refer to l_R. Then l wants to participate in the bloc if and only if

$$t_l(\pi_l^{1+} - \pi_l^{1-}) + (1-t_l)(\pi_l^{0+} - \pi_l^{0-}) \geq 0,$$

which, since $\lceil r_R(N_R - 1) \rceil = \lceil r_R N_R \rceil$, is equal to

$$t_l(P_l^{1+} - P_l^{1-}) + (1-t_l)(P_l^{0+} - P_l^{0-}) \geq 0,$$

Suppose $P_l^{1+} - P_l^{1-} \geq P_l^{0+} - P_l^{0-}$, then the expression is increasing in t_l and the cutoff that makes the agent indifferent is $\frac{P_l^{0-} - P_l^{0+}}{P_l^{1+} - P_l^{1-} + P_l^{0-} - P_l^{0+}}$. Hence, it suffices to show that $P_l^{1+} - P_l^{1-} \geq P_l^{0+} - P_l^{0-}$.

$$P_l^{1+} - P_l^{1-} = P\left[\sum_{i \in R_{-l}} p_i = \lceil r_R N_R \rceil - 1\right] P_1 - P\left[\sum_{i \in R_{-l}} p_i \leq \lfloor (1-r_R) N_R \rfloor - 1\right] P_3,$$

$$P_l^{0+} - P_l^{0-} = P\left[\sum_{i \in R_{-l}} p_i = \lfloor (1-r_R) N_R \rfloor\right] P_2 - P\left[\sum_{i \in R_{-l}} p_i \geq \lceil r_R N_R \rceil\right] P_4,$$

where P_1, P_2, P_3 and P_4 are as defined in the proof of lemma 17. As shown in the proof of claim 15, since R_{-lh_R} leans right, for any $r_R \geq \frac{N_R+1}{2N_R}$,

$$P\left[\sum_{i \in R_{-l}} p_i = \lceil r_R N_R \rceil - 1\right] \geq P\left[\sum_{i \in R_{-l}} p_i = \lfloor (1-r_R) N_R \rfloor\right]$$

and

$$P\left[\sum_{i \in R_{-l}} p_i \geq \lceil r_R N_R \rceil\right] \geq P\left[\sum_{i \in R_{-l}} p_i \leq \lfloor (1-r_R) N_R \rfloor - 1\right].$$

As shown in the proof of lemma 17, $P_1 \geq P_2$ and $P_4 \geq P_3$. Therefore, $P_l^{1+} - P_l^{1-} \geq P_l^{0+} - P_l^{0-}$.
■

Proof of Proposition 3

Proof. Let h_L be denoted in this proof by h. By proposition 2, it is a best response for every $i \in L$ to join if $\varepsilon(\pi_h^{1+} - \pi_h^{1-}) + (1-\varepsilon)(\pi_h^{0+} - \pi_h^{0-}) > 0$. Consider first the case $\lceil r_L N_L \rceil = \lceil r_L (N_L - 1) \rceil$. Let

$$\begin{aligned}
P_5 &\equiv P\left[\sum_{i \in M \sqcup R} v_i \in \left[\frac{N_M + N_R}{2} - \frac{N_L - 1}{2}, \frac{M + N_R}{2} + \frac{N_L - 1}{2} - \lceil r_L N_L \rceil\right]\right], \\
P_6 &\equiv P\left[\sum_{i \in M \sqcup R} v_i \in \left[\frac{N_M + N_R}{2} - \frac{N_L - 1}{2} + \lceil r_L N_L \rceil, \frac{M + N_R}{2} + \frac{N_L - 1}{2}\right]\right], \\
P_7 &\equiv \left[\sum_{i \in M \sqcup R} v_i = \frac{N_M + N_R}{2} + \frac{N_L - 1}{2}\right] \text{ and} \\
P_8 &\equiv P\left[\sum_{i \in M \sqcup R} v_i = \frac{N_M + N_R}{2} - \frac{N_L - 1}{2}\right].
\end{aligned}$$

Then,

$$\pi_h^{1+} - \pi_h^{1-} = P\left[\sum_{i \in L_{-h}} p_i = \lceil r_L N_L \rceil - 1\right] P_5 - P\left[\sum_{i \in R_{-l}} p_i \leq \lfloor (1-r_L) N_L \rfloor - 1\right] P_7$$

and

$$\pi_h^{0+} - \pi_h^{0-} = P\left[\sum_{i\in L_{-h}} p_i = \lfloor (1-r_L)N_L \rfloor\right] P_6 - P\left[\sum_{i\in L_{-h}} p_i \geq \lceil r_L N_L \rceil\right] P_8.$$

Let

$$\gamma = \frac{(N_L-1)!}{(\lceil r_L N_L \rceil - 1)!(N_L - \lceil r_L N_L \rceil)!}.$$

Then,

$$\varepsilon(\pi_h^{1+} - \pi_h^{1-}) + (1-\varepsilon)(\pi_h^{0+} - \pi_h^{0-}) < \varepsilon\gamma\varepsilon^{\lceil r_L N_L \rceil - 1} P_5 - \varepsilon(1-\varepsilon)^{N_L-1} P_7 + (1-\varepsilon)\gamma\varepsilon^{\lfloor (1-r_L)N_L \rfloor} P_6. \quad (10)$$

Note that $\lceil r_L N_L \rceil = \lceil r_L(N_L-1) \rceil$ and $r_L \leq \frac{N_L-2}{N_L-1}$ imply $r_L \leq \frac{N_L-2}{N_L}$. Divide the right-hand side of inequality 10 by ε, assume that $r_L \leq \frac{N_L-2}{N_L}$ and take the limit as ε goes to zero,

$$\lim_{\varepsilon\to 0} \gamma\varepsilon^{\lceil r_L N_L \rceil - 1} P_5 - (1-\varepsilon)^{N_L-1} P_7 + (1-\varepsilon)\gamma\varepsilon^{\lfloor (1-r_L)N_L \rfloor - 1} P_6 = -P_7 < 0.$$

Hence, if ε is low enough, $\varepsilon(\pi_h^{1+} - \pi_h^{1-}) + (1-\varepsilon)(\pi_h^{0+} - \pi_h^{0-}) < 0$ and it is not a best response for h to participate in the voting bloc.

Assume instead that $\lceil r_L(N_L-1) \rceil = \lceil r_L N_L \rceil - 1$. Let

$$\begin{aligned}
P_9 &\equiv P\left[\sum_{i\in M\sqcup R} v_i \in [\frac{N_M+N_R}{2} - \frac{N_L+1}{2} + \lceil r_L N_L \rceil, \frac{N_M+N_R}{2} + \frac{N_L-3}{2}\right], \\
P_{10} &\equiv P\left[\sum_{i\in M\sqcup R} v_i \in [\frac{N_M+N_R}{2} - \frac{N_L-3}{2}, \frac{N_M+N_R}{2} + \frac{N_L+1}{2} - \lceil r_L N_L \rceil\right].
\end{aligned}$$

Then,

$$\pi_h^{1+} - \pi_h^{1-} = P\left[\sum_{i\in L_{-h}} p_i = \lfloor (1-r_L)N_L \rfloor\right] P_9 - P\left[\sum_{i\in L_{-h}} p_i \leq \lfloor (1-r_L)N_L \rfloor - 1\right] P_7$$

and

$$\pi_h^{0+} - \pi_h^{0-} = P\left[\sum_{i\in L_{-h}} p_i = \lceil r_L N_L \rceil - 1\right] P_{10} - P\left[\sum_{i\in L_{-h}} p_i \geq \lceil r_L N_L \rceil\right] P_8.$$

and thus

$$\varepsilon(\pi_h^{1+} - \pi_h^{1-}) + (1-\varepsilon)(\pi_h^{0+} - \pi_h^{0-}) < \varepsilon\gamma\varepsilon^{\lfloor (1-r_R)N_R \rfloor} P_9 - \varepsilon(1-\varepsilon)^{N_L-1} P_7 + (1-\varepsilon)\gamma\varepsilon^{\lceil r_R N_R \rceil - 1} P_{10}. \quad (11)$$

Divide the right hand side by ε and take the limit as ε goes to zero.

$$\lim_{\varepsilon\to 0} \gamma\varepsilon^{\lfloor (1-r_R)N_R \rfloor} P_9 - (1-\varepsilon)^{N_L-1} P_7 + (1-\varepsilon)\gamma\varepsilon^{\lceil r_R N_R \rceil - 2} P_{10} = -P_7 < 0.$$

Hence, if ε is low enough, $\varepsilon(\pi_h^{1+} - \pi_h^{1-}) + (1-\varepsilon)(\pi_h^{0+} - \pi_h^{0-}) < 0$ and it is not a best response for h to participate in the voting bloc. ■

Proof of Proposition 5

Proof. By construction. I first show that if $\hat{r}_R = \frac{N_R-1}{N_R}$ and $t_i = t_R \geq \frac{1}{2}$ for any $i \in R$, then $a_i = 1$ is a mutual best response for all $i \in R$. Note that the voting bloc $(R, \hat{r}_R)$ only affects the outcome in the assembly if $\sum_{i \in L \sqcup M} v_i = \frac{N+1}{2} - N_R$ and $\sum_{i \in R} p_i = N_R - 1$, or $\sum_{i \in L \sqcup M} v_i = \frac{N+1}{2} - 1$ and $\sum_{i \in R} p_i = 1$. Given the assumptions on sizes and distributions of types, these events occur with positive probability. In either of these events, making the minority of one in R vote with the majority of R reverses the outcome, to the detriment of one agent in R, and to the benefit of $(N_R - 1)$ members of R. generating a net aggregate gain of $N_R - 2$ to the members of bloc R. Since all members of R are ex-ante identical, the ex-ante expected surplus for each member of R, relative to the benchmark in which the members of R do not coordinate their votes is strictly positive. If agent i leaves the bloc, then $\frac{N_R-2}{N_R-1} < \hat{r}_R$, so the members of the diminished voting bloc $(R \backslash i, r_R)$ only achieve a sufficiently large internal majority if they all agree in their preferences, so i's departure makes members of R not coordinate their actions. But i's utility is higher if she belongs to the bloc than if members of R do not coordinate their votes, so it follows that i strictly prefers not to leave the bloc $(R, \hat{r}_R)$.

Since the utility function of every agent is continuous in the type of every agent, if we relax the assumption that $t_i = t_R$, and we assume instead that $t_i \in (t_R - \varepsilon, t_R + \varepsilon)$ $\forall i \in R$, then the supremum of the change in the utility for an agent that can occur as a result of this relaxation is continuous in ε for each agent. Thus, for a small enough ε, all agents strictly prefer not to leave the bloc $(R, \hat{r}_R)$.

It now suffices to show that if r_R is simple majority, it is not a mutual best response for all $i \in R$ to join the bloc. Let $P\left[\sum_{i \in L \sqcup M} v_i = \frac{N+1}{2} - N_R\right] = \lambda$. By the assumption on sizes and types of M and L, $\lambda > 0$. Let $t_R = 1 - \delta$. Let E be the event that $i \in R$ rejects the proposal, a majority of R favors the proposal, and $\sum_{i \in L \sqcup M} v_i = \frac{N+1}{2} - N_R$. In this event, i is better off ex-post if she is not part of the bloc. Note that

$$\lim_{\varepsilon \to 0} P[E] = \delta \lambda \sum_{k=\frac{N_R+1}{2}}^{N_R-1} Bi(N_R - 1, 1 - \delta; k),$$

where $Bi(n, p; k)$ denotes the probability that a binomial distribution with parameters (n, p) takes a value of k. Then,

$$\lim_{\delta \to 0} \frac{1}{\delta} \left(\lim_{\varepsilon \to 0} P[E] \right) = \lambda > 0,$$

so that for any $\bar{\lambda} \in (0, \lambda)$, there exist $\bar{\delta}_{\bar{\lambda}} > 0$ and $\bar{\varepsilon}_{\bar{\lambda},\delta} > 0$ such that for any $\delta \in (0, \bar{\delta})$, and any $\varepsilon < \bar{\varepsilon}_{\bar{\lambda},\delta}$,

$$\frac{1}{\delta} P[E] > \bar{\lambda} > 0.$$

Agent i is better off inside the bloc only if the rest of the bloc is tied. But

$$\lim_{\delta \to 0} \frac{1}{\delta} \left(\lim_{\varepsilon \to 0} P \left[\sum_{i \in R} p_i = \frac{N_R - 1}{2} \right] \right) = 0.$$

Therefore, for a sufficiently low δ, there exists a sufficiently low ε such that the probability that i is better off ex-post outside the bloc outweighs the probability that i is better off inside the bloc, and ex-ante i prefers to leave the bloc. ■

Proof of Proposition 7

Proof. Let C denote C_R. Let $r_R(C, p_C)$ denote the rule given that subset $C \subseteq R$ with preferences p_C accept party discipline. Let $r_R(C, p_C) = p_C$ for any C such that $|C| = N_R - 1$ and for any p_C. That is, the party does not coordinate votes if exactly one member rejects party discipline. Let $r_R(R, p_R) = p_R$ for any p_R such that $\sum\limits_{i \in R} p_i \neq N_R - 1$. That is, the party does not coordinate votes unless exactly $N_R - 1$ members favor the proposal. For any $i \in R$, let

$$\gamma_i \equiv (1 - t_i) \Pr[g^{R \backslash \{i\}}(N_R - 1) | p_i = 0] \Pr \left[\sum_{i \in L \sqcup M} v_i = \frac{N+1}{2} - N_R | p_i = 0, p_j = 1 \forall j \in R \backslash \{i\} \right].$$

That is, γ_i is the probability that i is against the proposal, everyone else in R is in favor, and the proposal falls one vote short of passage if R does not coordinate votes. Let $\hat{\gamma} \equiv \min\limits_{i \in R} \gamma_i$. Note that since $\Omega(p) > 0$, for any $i \in R$, the probability of the event in which i opposes the proposal, every other agent in R favors it, and every agent in L opposes it is strictly positive. Since r_L is weakly Pareto optimal for any $C_L \subseteq L$, then the sum of votes by members of L for the proposal is zero in this event and $\sum_{i \in L \sqcup M} v_i = \sum_{i \in M} v_i$. Since $N_L, N_R \leq \frac{N-1}{2}$,

$$\frac{N+1}{2} - N_R = \frac{N_M + N_L - N_R + 1}{2} \leq N_M,$$

and thus $\gamma_i > 0$, so that $\hat{\gamma} > 0$ as well.

For any $i \in R$ and any p_R such that $p_i = 0$ and $p_j = 1$ $\forall j \in R \backslash \{i\}$, let

$$r_R(R, p_R) = \left\{ \begin{array}{c} (1, 1, ..., 1) \text{ with probability } \frac{\hat{\gamma}}{\gamma_i} \\ p_R \text{ with probability } 1 - \frac{\hat{\gamma}}{\gamma_i} \end{array} \right\}.$$

Then, under rule r_R, if $a_i = 1$ for every $i \in R$, then for any $i \in R$, the probability that party discipline in R results in $\{p_i \neq v_i = 1$ and $p_j = v_j = 1$ for any $j \in R \backslash \{i\}\}$ and that the outcome in the assembly changes from p_i to v_i as a result of party discipline is $\hat{\gamma}$. Thus, the probability that agent i is hurt by the coordination of votes is $\hat{\gamma} > 0$ while the probability that i benefits is $(N_R - 1)\hat{\gamma}$, for a net benefit of $(N_R - 2)\hat{\gamma}$. If i rejects party discipline, the bloc never coordinates under rule r_R, thus, i is strictly better off accepting party discipline under r_R. ■

Proof of Proposition 8

Proof. Given an arbitrary $i \in L \sqcup R$, assume that every $j \in L \sqcup R$ s.t. $j \neq i$ joins her voting bloc and follows the recommendation in the assembly, and assume every independent votes sincerely in the assembly. By the same argument as in lemma 1, $v_j = p_j$ is a best response for each $j \in M$. I then want to show that joining her bloc and following the recommendation is a best response for i. Without loss of generality, assume $i \in L$. For any p such that $d_L = 0$, $v_i = p_i$ is a best response, since by deviating to $v_i = 1 - p_i$, agent i either has no effect, or changes the outcome to $1 - p_i$, and further, i incurs a cost c in doing so. For any p such that $d_L = 1$, i's vote does not affect the outcome, hence following the recommendation is strictly better, since it avoids the cost c for i. Thus, at the voting stage in the assembly, every agent best responds by following her bloc's recommendation, or voting sincerely if the agent belongs to no bloc.

The statement of the proposition makes rules and recommendations depend on the true preference profile p, and not on the reported preferences $\widehat{p}_C$ at the internal meeting of the set of agents C, thus there is no strategic component at the internal meetings.

Finally, I must show that $a_i = 1$ is a best response for every agent. Note that with the proposed strategy profile, agents never pay cost c, and they never vote against their preference when they are individually pivotal. Hence they cannot become better off by deviating not to join. Thus the proposed strategy is indeed an equilibrium. To show that agents are strictly better off joining, it suffices to note that since $\Omega(p) > 0$ and $N_L, N_R \leq \frac{N-1}{2}$ with positive probability they are pivotal inside their bloc, and their bloc is pivotal in the assembly, so that if i does not join the bloc, the outcome is $1 - p_i$, but if i joins the bloc, the outcome in the assembly is p_i. ■

Note that if we change the assumptions of the proposition to let the voting rule r_J and the recommendation rule d_J depend on $\hat{p}_J$ and not on p_J for each $J \in \{L, R\}$, it is easily verified that lemma 1 still applies, and $\widehat{p_i} = p_i$ is a best response for each $i \in L \sqcup R$. Assume, without loss of generality, that $i \in L$ and $p_i = 1$. Reporting $\widehat{p_i} = 0$ instead either has no effect on $\sum_{i \in L} r_L^i(\hat{p}_L)$, or it reduces it, so either it has no effect on

$$\sum_{i \in L} r_L^i(\hat{p}_L) + \sum_{i \in R} r_R^i(\hat{p}_R) + \sum_{i \in M} p_i \tag{12}$$

or it reduces it. Reducing expression 12 cannot bring any advantage to i: It can only lower it from the set of values where voting is disciplined and the policy proposal passes (as preferred by i) to the set of values where agents vote according to true preferences; or from the set of values where agents vote according to true preferences, to a set of values where voting is disciplined and the proposal is rejected. Thus, it is without loss that I identify reports $\widehat{p}_i$ with true preference p_i in the statement of the proposition.

5.1 Proof of Proposition 9

Proof. For the equilibrium with party discipline, given rule r_V, let the equilibrium strategy s_i of each agent i be such that $a_i = V$ and $\widehat{p}_i(a, p_i) = p_i$ $\forall i \in \mathcal{N}$. Let $i \in \mathcal{N}$ be an arbitrary agent. For any realization of preferences p such that the equilibrium outcome is equal to p_i, agent i achieves her preferred outcome, so deviating she does not gain anything ex-post. For any realization of preferences p such that $v_i = p_i$ but the outcome is not equal to p_i, the number of other agents voting with i inside the bloc is at most $(N-2)r_V$, so if i leaves, these agents do not achieve the threshold of votes $(N-1)r_V$ to coordinate votes in the reduced bloc after $i's$ defection; consequently, the number of votes cast in the assembly according to $i's$ preference is at most the same as before $i's$ departure, so i gains nothing deviating. Finally, for any realization of preferences p such that $v_i \neq p_i$, agent i gains one vote (her own) if she deviates, but since i loses the internal vote as a member of the bloc, it follows that the internal vote in the reduced bloc after $i's$ departure also results in the bloc coordinating to vote against $i's$ preference. Hence, in the assembly, all other agents vote against $i's$ preference and, given that $r_{\mathcal{N}}$ is not unanimity, the outcome is not the policy preferred by i. Therefore, a deviation does not earn i a higher ex-post utility for any realization of preferences, so ex-ante, she does not benefit from a deviation, and the proposed strategy is indeed an equilibrium. Since Ω has full support, $p = (0, 1, 1, ..., 1)$ occurs with positive probability and, since $r_V \leq \frac{N-1}{N}$, if p occurs, $p_1 = 0$ but $\sum_{i \in \mathcal{N}} p_i = N - 1 \geq N r_V$, so $v_i = 1$ for every member i of the bloc $V = (\mathcal{N}, r_V)$, including i. Hence there is party discipline in equilibrium.

Note that if $r_V \leq r_{\mathcal{N}} - \frac{1}{N}$ and $\sum_{i \in \mathcal{N}} p_i \in [N r_V, N r_V + 1)$, then the proposal passes if the whole coalition forms voting bloc V, and it fails if there are no blocs. Since Ω has full support, p occurs with positive probability. Therefore, if $r_{\mathcal{N}} > \frac{N+1}{2N}$, the voting bloc $V = (\mathcal{N}, r_V)$ with $r_V = \frac{N+1}{2N}$ corresponds to a Nash equilibrium with relevant party discipline, in which $a_i = V$ and $\widehat{p}_i(a, p_i) = p_i$ $\forall i \in \mathcal{N}$. ■

Proof of Proposition 11

Proof. By contradiction. Suppose (absurd) that in equilibrium $\exists (C_j, r_j)$ such that r_j is simple majority and $\frac{N+1}{2} \leq N_j$. If $N_j = N$, then

$$\sum_{i \in \mathcal{N}} v_i \geq r_{\mathcal{N}} N \iff \sum_{i \in \mathcal{C}_j} v_i \geq \frac{N+1}{2} \iff \sum_{i \in \mathcal{C}_j} p_i \geq \frac{N+1}{2} \iff \sum_{i \in \mathcal{N}} p_i \geq r_{\mathcal{N}} N$$

so there is no relevant party discipline, a contradiction.

Suppose $N_j < N$. For any p such that $\sum\limits_{h\in C_j} p_h \neq \frac{N_j}{2}$, it follows that $\sum\limits_{h\in C_j} v_h \in \{0, N_j\}$ and the outcome coincides with the vote of the bloc; since the outcome is independent of the votes of outsiders, any $i \notin C_j$ is at least equally well off entering the voting bloc. For any p such that $\sum\limits_{h\in C_j} p_h = \frac{N_j}{2}$, any $i \notin C_j$ who joins the bloc causes $\sum\limits_{h\in C_j \sqcup i} v_h = p_i N_j$ and i wins in the assembly; if i was winning outside of the bloc, i is indifferent between entering or remaining an outsider, but if i was losing, i is strictly better off entering the bloc.

If the bloc is odd sized, i is strictly better off entering the bloc for p such that $\sum\limits_{h\in C_j} p_h = \sum\limits_{h\in \mathcal{N}} p_h = \frac{N_j+1}{2}$. Then $\sum\limits_{h\in C_j} v_h = \sum\limits_{h\in \mathcal{N}} v_h = N_j \geq \frac{N+1}{2}$ and the proposal passes. If $i \notin C_j$ joins the bloc, then the expanded bloc is tied, $\sum\limits_{h\in \mathcal{N}} v_h = \sum\limits_{h\in C_j \sqcup i} v_h = \sum\limits_{h\in C_j} p_h < \frac{N+1}{2}$ and the proposal fails. Therefore, regardless of the size of the bloc, for any $i \notin C_j$ there exist a preference profile for which i is strictly better off joining the bloc. Since Ω has full support, every preference profile occurs with positive probability and every non-member strictly prefers to join the bloc. Summarizing, $a_i \neq j$ is not a best response for any $i \in \mathcal{N}$, which contradicts (C_j, r_j) with $N_j < N$ forming in equilibrium.

Suppose (absurd) that in equilibrium $a_i = 0$ for some $i \in \mathcal{N}$, and $\exists (C_k, r_k)$ with $r_k \leq \frac{N-1}{2N_K}$ and $N_k \geq \frac{N+1}{2}$. Let s' be a new strategy profile such that $a_i' = k$ and all else is unchanged. For any preference profile p such that $v_i = p_i$ under the strategy profile s', the ex-post utility for i under s' and p is equal or greater than the utility under s and p, since i joining the bloc can never reduce the number of votes cast by other bloc members for the option preferred by i. For any profile p' such that $v_i \neq p_i'$ under strategies s', the bloc votes against i in the assembly both if i joins and if she does not; since the bloc acting together with or without i is a dictator in the assembly, with preferences p', the ex-post utility of i is zero under s and s', so i is in this case indifferent. There is no case in which i is worse off joining the bloc. Let p'' be such that $\sum\limits_{h\in C_k} p_h'' = (1-r_k)N_k$ and $p_l'' = 1$ for all $l \notin C_k$. With strategy profile s and preferences p'', $\sum\limits_{h\in C_k} v_h = 0$ and the proposal is rejected. With strategy profile s',

$$\begin{aligned}
\sum_{h\in C_k \sqcup i} p_h'' &= (1-r_k)N_k + 1 > (1-r_k)(N_k+1) \text{ and } \sum_{h\in C_k \sqcup i} v_h \geq \sum_{h\in C_k \sqcup i} p_h'' = (1-r_k)N_k + 1 \\
\text{so } \sum_{h\in \mathcal{N}} v_h &\geq (1-r_k)N_k + N - N_k = N - r_k N_k \geq N - \frac{N-1}{2N_K} N_k = \frac{N+1}{2},
\end{aligned}$$

the policy proposal passes in the division of the assembly and i is better off -a contradiction. ■

Proof of Proposition 12

Proof. By construction. I show that with sufficient polarization, the formation of two blocs of odd size n with simple majority is supported in equilibrium.

Let Ω satisfy Condition 1. Let $r_1 = r_2 = \frac{N+1}{2N}$. For a given λ, let n_λ denote the largest odd number smaller than $n_K + 1$ such that if n_λ agents with type t_{K-} form a bloc with rule r_1, n_λ agents with type t_{K+} form a bloc (C_2, r_2) and every other agent remains independent, then the agents in each of the two blocs strictly prefer to remain in their bloc than to become independents. Since agents with type t_{K-} are identical, they all share equally the benefits of forming a bloc, so if only three form a bloc and a single defection disbands the bloc, none of the three has an incentive to disband the bloc to become an independent (see the proof of proposition 5). The same logic applies to the agents with type t_{K+}. Therefore, $n_\lambda \geq 3$. Since the members of each bloc strictly benefit, and utilities are continuous in the probability distribution over preference profiles, there exists some open neighborhood of Ω such that the same agents (now with not exactly the same types) benefit strictly from participating in these blocs.

It remains to be shown that no agent wants to join a bloc. It suffices to show that no agent wishes to join the bloc $V_1 = (C_1, r_1)$ of size n_λ. The proof is similar to the proof of proposition 3. Let $\varepsilon_\lambda = \frac{1}{2\gamma_K} - \lambda$ and let $O_\lambda(\Omega)$ be the open neighborhood of radius λ around Ω. Given a probability distribution over preference profiles $\Omega' \in O_\lambda(\Omega)$, for an agent $i \notin C$, the net benefit of joining the bloc is:

$$P[p_i=1]\left\{\begin{array}{c} P\left[\sum_{j\in C_1} p_j = \frac{n_\lambda-1}{2} | P[p_i=1]\right] P\left[\frac{N+1}{2N} - \frac{n_\lambda+1}{2} \leq \sum_{j\in\mathcal{N}\setminus\{C_1\sqcup\{i\}\}} v_j \leq \frac{N-3}{2N} | P[p_i=1]\right] \\ -P\left[\sum_{j\in C_1} p_j < \frac{n_\lambda-1}{2} | P[p_i=1]\right] P\left[\sum_{j\in\mathcal{N}\setminus\{C_1\sqcup\{i\}\}} v_j = \frac{N-1}{2N} | P[p_i=1]\right] \end{array}\right\} +$$
$$P[p_i=0]\left\{\begin{array}{c} P\left[\sum_{j\in C_1} p_j = \frac{n_\lambda+1}{2} | P[p_i=0]\right] P\left[\frac{N+1}{2N} - n_\gamma \leq \sum_{j\in\mathcal{N}\setminus\{C_1\sqcup\{i\}\}} v_j \leq \frac{N-1}{2N} - \frac{n_\lambda+1}{2} | P[p_i=0]\right] \\ -P\left[\sum_{j\in C_1} p_j > \frac{n_\lambda+1}{2} | P[p_i=0]\right] P\left[\sum_{j\in\mathcal{N}\setminus\{C_1\sqcup\{i\}\}} v_j = \frac{N-1}{2N} - n_\lambda | P[p_i=0]\right] \end{array}\right\}.$$

The limit of this expression divided by $P[p_i = 1]$ as $\lambda \to \frac{1}{2\gamma_K}$ is equal to:

$$\lim_{\lambda\to\frac{1}{2\gamma_K}} -P\left[\sum_{j\in\mathcal{N}\setminus\{C_1\sqcup\{i\}\}} v_j = \frac{N-1}{2N}\right] = \left\{\begin{array}{c} \lim_{\lambda\to\frac{1}{2\gamma_K}} -P\left[\sum_{j\in\mathcal{N}\setminus\{C_1\sqcup C_2\sqcup i\}} p_j = \frac{N-1-n_\lambda}{2N}\right] < 0 \text{ if } i \notin C_2. \\ \lim_{\lambda\to\frac{1}{2\gamma_K}} -P\left[\sum_{j\in\mathcal{N}\setminus\{C_1\sqcup C_2\}} p_j = \frac{N-n_\lambda}{2N}\right] < 0 \text{ if } i \in C_2. \end{array}\right\}$$

Hence in the limit, i does not want to join the voting bloc V_1 and the profile with two blocs with simple majority, both of size $n_\lambda \geq 3$ and sincere voting is a Nash equilibrium with relevant party discipline. ■

5.2 Proof of Claim 13

Proof. I show that given the probability distribution Ω_ε, there is an equilibrium with $2K$ blocs in which each agent is strictly better off than if she deviates. Therefore, by continuity of the payoff functions, the same equilibrium holds in an open neighborhood around Ω_ε.

Given Ω_ε, blocs are either identical or symmetric to each other. Therefore, it suffices to show that agent 1 has no incentive to deviate to become an independent, to join agents $\{4,5,6\}$ or to join $\{6K-2, 6K-1, 6K\}$, and that agent N has no incentive to join $\{1,2,3\}$.

A deviation to become independent by a member of a bloc can only make the deviator worse off, because members of a bloc are ex-ante identical, hence they benefit equally from the net gains of the bloc, and since blocs are of size three, they cease to coordinate if an agent leaves. Since Ω_ε has full support, the gains lost if the bloc does not coordinate are strictly positive.

Note that as $\varepsilon \longrightarrow 0$, the probability that agent $i \in \{1, N\}$ is worse if she deviates to join $\{6K-2, 6K-1, 6K\}$ converges to $\frac{1}{2}$, whereas the probability that she is better off converges to zero. So the only remaining deviation to explore is for agent 1 to join $\{4,5,6\}$. The event in which i is strictly better off after this deviation is $E_1 \cup E_2$, where E_1 and E_2 are defined as follows.

E_1 : The realized preference profile p is such that $p_2 = p_3 \neq p_1$, $|\{j \in \{4,5,6\} : p_j = p_i\}| \geq 1$, $p_N \neq p_1$ and a majority in exactly K blocs has the same preference as agent 1, and

E_2 : The realized preference profile p is such that $p_2 = p_3 = p_1$, $|\{j \in \{4,5,6\} : p_j = p_i\}| = 1$, $p_N \neq p_1$ and a majority in exactly K blocs has the same preference as agent 1.

Note that $\lim_{\varepsilon \longrightarrow 0} P[E_1] = \lim_{\varepsilon \longrightarrow 0} P[E_2] = 0$. In particular, $P[E_1]$ is of order ε^2, since it requires at least two agents (i and an agent in $\{4,5,6\}$ suffices) to have the opposite preference from their ex-ante most probable preference. $P[E_2]$ is of order ε^4; as it requires that at least four agents have the opposite preference to their ex-ante most probable one (for instance, agents $\{4, 5, 6K-1, 6K\}$).

The event in which i is strictly worse off after the deviation is $E_3 \cup E_4$, where E_3 and E_4 are defined as follows.

E_3 : The realized preference profile p is such that $p_2 = p_3 = p_1$, $|\{j \in \{4,5,6\} : p_j = p_i\}| = 0$, $p_N = p_1$ and a majority in exactly K blocs has the same preference as agent 1.

E_4 : The realized preference profile p is such that $p_2 \neq p_3$, $|\{j \in \{4,5,6\} : p_j = p_i\}| \neq 1$, $p_N = p_1$ and a majority in exactly K blocs has the same preference as agent 1.

Event E_3 is of order ε^5, as it requires all three members of one of the blocs that was expected to oppose the proposal, and a majority of a bloc that was expected to favor the proposal, to all have the opposite preferences. Event E_4 is of order ε, as it suffices that either agent 2 or agent

3 favor the proposal, while every other agent has their most probable preference. Thus, if ε is sufficiently small, the probability of event E_4 dominates all others, and agents strictly prefer not to deviate. ■

Proof of Proposition 14

Let (π, r) be such that $\pi_0 = \{1, 2\}$, and $\pi_1 = \{3, 4, ..., N\}$ with $r_1 = \frac{N-1}{2N-4}$. If $i \geq 3$ leaves the bloc, the bloc remains a dictator, so i cannot achieve a better policy outcome than staying in the bloc. It follows that no member wishes to leave the bloc. Suppose instead that agent 2 enters the bloc. Entrance only affects the outcome if $\sum_{i>1} p_i = \frac{N-1}{2}$ and $p_1 = p_2$ so the outcome coincides with the preference of the new entrant, even though without agent 2, a majority of one inside the bloc was against the preference of agent 2. The entrance of agent 2 generates a net loss in the sum of ex-post utility for former members of the bloc in any event in which it has an effect on the outcome. Since Ω has full support, it then generates a strict loss of in the sum of ex-ante utility. It must then be than ex-ante at least one of the members of the bloc is worse off if 2 enters, so 2 cannot enter with $\alpha \geq 1$.

References

Aldrich, John H. 1995. *Why Parties? The Origin and Transformation of Political Parties in America*. Chicago, University of Chicago Press.

Aldrich, John H & David W Rohde. 2001. *Congress Reconsidered, 7th Ed.* CQ Press chapter The Logic of Conditional Party Government: Revisiting the Electoral Connection.

Baron, David. 1993. "Government Formation and Endogenous Parties." *American Political Science Review* 85:137–164.

Baron, David P. 1989. "A Noncooperative Theory of Legislative Coalitions." *American Journal of Political Science* 33(4):1048–1084.

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

Bloch, Francis & Armando Gomes. 2006. "Contracting with Externalities an Outside Options." *Journal of Economic Theory* 127(1):171–201.

Buchanan, James M. & Gordon Tullock. 1962. *The Calculus of Consent*. Ann Arbor, University of Michigan Press.

Caillaud, Bernard & Jean Tirole. 1999. "Party Governance and Ideological Bias." *European Economic Review* 34:779–789.

Caillaud, Bernard & Jean Tirole. 2002. "Parties as Political Intermediaries." *Quarterly Journal of Economics* 117(4):1453–1489.

Calvert, Randall L. & Nathan Dietz. 2005. *Social Choice and Strategic Decisions*. Springer-Verlag chapter Legislative Coalitions in a Bargaining Model with Externalities.

Caro, Robert A. 2002. *The Years of Lyndon Johnson: Master of the Senate*. Alfred A. Knopf.

Carraro, Carlo. 2003. *The Endogenous Formation of Economic Coalitions*. Edward Elgar Publishing.

Cox, Gary W. & Matthew D. McCubbins. 1993. *Legislative Leviathan*. Berkeley, University of California Press.

Darroch, John N. 1964. "On the distribution of the Number of Successes in Independent Trials." *The Annals of Mathematical Statistics* 35(3):1317–1321.

Diermeier, Daniel & Razvan Vlaicu. 2008. "Parties, Coalitions, and the Internal Organization of Legislatures." SSRN wp 1150963.

Drèze, Jacque & Joseph Greenberg. 1980. "Hedonic Coalitions: Optimality and Stability." *Econometrica* 48(4):987–1003.

Duverger, Maurice. 1959. *Political Parties*. London, Methuen.

Eguia, Jon X. forthcoming. "Endogenous Parties in an Assembly." *American Journal of Political Science* pp. –.

Felsenthal, Dan S & Moshé Machover. 2002. "Annexations and Alliances: When are Blocs Advantageous a priori." *Social Choice and Welfare* 19:295–312.

Fox, Justin. 2006. "Legislative Cooperation among Impatient Legislators." *Journal of Theoretical Politics* 18(1):68–97.

Gelman, Andrew. 2003. "Forming Voting Blocs and Coalitions as Prisoner's Dilemma: A Possible Theoretical Explanation for Political Instability." *Contributions to Economic Analysis and Policy* 2(1).

Hart, Sergiu & Mordechai Kurz. 1983. "Endogenous Formation of Coalitions." *Econometrica* 51(1):1047–1064.

Hix, Simon, Abdul Noury & Gérard Roland. 2007. *Democratic Politics in the European Parliament*. Cambridge University Press.

Hosli, Madeleine O. 2000. "Coalition Formation and Voting in the Council of the European Union." *Homo Oeconomicus* 17:31–55.

Hosli, Madeleine O. 2002. "Preferences and Power in the European Union." *Homo Oeconomicus* 19(3):311–326.

Humphreys, Macartan. 2008. "Theory of Coalitions and Its Applications." *Annual Review of Political Science* 11:351–386.

Hyndman, Kyle & Debraj Ray. 2007. "Coalition Formation with Binding Agreements." *Review of Economic Studies* 74(4):1125–1147.

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

Laruelle, Annick & Federico Valenciano. 2005. "Assessing Success and Decisiveness in Voting Situations." *Social Choice and Welfare* 24:171–197.

Levy, Gilat. 2004. "A Model of Political Parties." *Journal of Economic Theory* 115(2):250–277.

Maggi, Giovanni & Massimo Morelli. 2006. "Self Enforcing Voting in International Organizations." *American Economic Review* 96(4):1137–1158.

Morelli, Massimo. 2004. "Party Formation and Policy Outcomes under Different Electoral Systems." *The Review of Economic Studies* 71(3):829–853.

Osborne, Martin & Rabee Tourky. 2008. "Party Formation in Single Issue Politics." *Journal of the European Economic Association* 6(5):974–1005.

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

Rohde, David W. 1991. *Parties and Leaders in the Postreform House*. Unversity of Chicago Press.

Snyder, James & Michael Ting. 2002. "An Informational Rationale for Political Parties." *American Journal of Political Science* 46:90–110.

Thrall, Robert M. & William F Lucas. 1963. "N-person Games in Partition Form." *Naval Research Logistics Quarterly* 10:281–298.

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