Endogenous Parties in an Assembly.*

Jon X. Eguia[†]

New York University

February 22, 2010

*I am grateful to Justin Fox, Sandy Gordon, Dimitri Landa, Jeffrey Lax, Matt Jackson, Andrew Martin, Tom Palfrey and Keith Poole for their suggestions and their help.

[†]Assistant Professor of Politics, New York University, 19 West 4th Street, New York, NY 10012 (eguia@nyu.edu).

Abstract

In this paper I show how members of an assembly form voting blocs strategically to coordinate their votes and affect the policy outcome chosen by the assembly. In a repeated voting game, permanent voting blocs form in equilibrium. These permanent voting blocs act as endogenous political parties that exercise party discipline. In a stylized assembly I prove that the equilibrium parties must be two small polarized voting blocs, one at each side of the ideological divide.

Keywords: Party formation. Voting blocs. Party discipline. Endogenous parties.

Members of a democratic assembly that chooses a policy outcome by means of voting can affect the policy outcome chosen by the assembly by coordinating their voting behavior and forming a *voting bloc*: that 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. The most prominent examples of voting blocs are disciplined political parties in legislative assemblies and parliaments; indeed, in most national legislatures, legislators of both the ruling and opposition party frequently vote along party lines (Carey, 2007).

Who benefits when agents with diverse preferences coordinate their votes and form a voting bloc? What voting blocs do we expect to find in an assembly with heterogeneous members? How and why do voting blocs change over time? Why do legislators form parties and vote together? Why do we so frequently find more than one party in an assembly? In this article I address these and other questions by studying an assembly with a finite number of agents who can coordinate with each other to form voting blocs before voting on a policy proposal.

I explain the emergence of multiple political parties as competing coalitions of agents who coordinate their votes. A group of members of an assembly –a party- strategically coalesce into a voting bloc to coordinate their votes, seeking to influence the policy outcome for an ideological gain. Party members commit to accept the party discipline and to vote for the party line, which is chosen according to an aggregation rule internal to the party.

I consider an assembly in which any collection of subsets of agents can coalesce into voting blocs. I show that in equilibrium, voting blocs with strict party discipline form to affect voting records and policy outcomes. I prove that voting blocs of size less than minimal winning can generate an ideological policy gain to their members even though they are not big enough to guarantee a majority; it follows that there are equilibria in which different groups of agents simultaneously form competing voting blocs. To obtain sharper predictions about the voting blocs that form, I apply the theory to a stylized assembly with 9 members whose preferences are symmetrically distributed. Using a solution concept that allows for coalitional deviations in which at most one bloc splits apart, I find that if the assembly is sufficiently polarized, then two voting blocs form, one at each side of the ideological spectrum with a group of independents including the median in between the two blocs; in other words, the theory predicts a two party system emerging in the assembly.[1]

These predictions and the theory that generates them are based only on the properties of majority voting as a rule to aggregate preferences; they do not rely on campaigns and elections, or any other element beside the act of voting in an assembly. While getting candidates elected is one of the major activities of political parties, Duverger (1959) notes that most political parties in Western democracies come to exist when legislators serving in an assembly coalesce in factions to further their policy goals, and these factions start coordinating electoral campaigns only at a later stage of their development. Duverger (1959) refers to these parties that originate as coalitions within an assembly as parties of "intraparliamentary origin."[2]

Within the literature on the formation of parties of intraparliamentary origin, we can distinguish two streams. The first approach sees the majority party as a cartel that gains monopolistic control over the agenda of policy proposals that are considered by the assembly

[1] In a separate working paper (Eguia, 2010), I develop the theory further, studying the formation of voting blocs within each of two exogenous parties, and the comparative statics on the formation of endogenous parties with respect to the types of the agents and the voting rule used in the assembly.

[2] For complementary theories of party formation based on the role of parties as organizations that run electoral campaigns and help candidates get elected for office, see Snyder and Ting (2002), Levy (2004), Morelli (2004), or Ashworth and Bueno de Mesquita (2008).

(Cox and McCubbins, 1993; Cox and McCubbins, 2007; Diermeier and Vlaicu, 2008). These agenda-based theories of party formation fit the institutional framework of the US Congress and its powerful committees particularly well.

The second approach describes parties as voting coalitions: sets of legislators who agree to coordinate or trade their votes. Authors advocating this approach have shown that members of a winning coalition of legislators have incentives to coordinate a logroll or trade of votes whenever this trade makes all the legislators who join the logroll better off, and similarly, that a winning coalition of legislator have incentives to form a party to allocate the pork available for distribution among party members only (Schwartz, 1989; Baron, 1993; Aldrich, 1995; Carruba and Volden, 2000; Jackson and Moselle, 2002). Furthermore, a majority coalition of legislators also have incentives to cooperate by trading votes in an infinitely repeated game to change ideological policies (Fox, 2006).

The agenda-based theories convincingly highlight the incentives to form or join a majority party, but they cannot provide an equally persuasive rationale for forming or joining an opposition party is unable to influence the agenda. The theories of parties as voting coalitions have also focused on the formation of a single party that is able to command a winning majority in an assembly. Yet, there is some empirical evidence that is prima facie at odds with these explanations: Party cohesion and unity in voting is as good in smaller parties as in the largest one in each assembly, and in presidential systems, the opposition party is as united and cohesive as the ruling party (Carey, 2007). The puzzle that remains unexplained is the emergence of a party system in which two or more parties form and oppose each other, even though no minority party can control the agenda or win votes to distribute benefits or change policy on its own.

Noticing that voting blocs do not need to be able to win on their own to be effective solves this puzzle. If the majority party is able to coordinate only imperfectly (Rohde, 1991; Aldrich and Rohde, 2001), or if the largest party does not have a majority size, then a minority party can influence the outcome by coordinating its own votes and picking off a few stray votes from the rest of the assembly.

In a classic work, Schattschneider (1942) recognizes this incentive to form opposing parties, and notes that organization by one faction leads to counterorganization by an opposing faction, so that the coordination of votes within each of these two factions leads to the emergence of a two party system. I substantiate this basic insight by providing systematic micro-foundations that account for it, and building on these foundations, I develop a theory of parties as voting blocs that characterizes the incentives that each arbitrary set of agents face to coordinate with each other, so that given the characteristics of the agents, we can predict how many parties form, and which agents join each of them.

Motivating Examples

In this subsection I present two examples to illustrate how the formation of voting blocs affects voting results and policy outcomes.

Example 1 *Let there be an assembly with five agents A, B, C, D, E who have to make a binary choice decision -to approve or reject some action- by simple majority. Suppose that only agents C, D favor the proposal, so if agents vote their individual preference, the proposal is rejected 2-3. If agents B, C, D form a voting bloc that commits to vote together according to the preferences of the majority of members of the bloc, C, D in favor of the action achieve* *an internal majority and with the three votes of the bloc B, C, D the action is approved. This outcome makes B worse off so in this case B has no incentives to join such a bloc. However, suppose instead that there are three different actions to be approved or rejected on three different topics. Suppose further that only C, D favor action one, only B, C favor action two and only B, D favor action three. Then, without voting blocs, all actions are rejected 2-3. Agents B, C and D get their desired outcome only in one decision. If they form a voting bloc, the majority of the bloc favors all three actions, and all three are approved. Agents B, C and D are all better off, achieving their desired outcome in two decisions.*

Each member of the voting bloc benefits from joining in. Each agent is forced to vote against her wishes in one instance, but more often (twice), belonging to the bloc allows the agent to gather enough votes to sway the decision of the assembly to her preferred outcome. The same incentive to join a voting bloc exists if agents vote over a single decision, but with some uncertainty over preferences. For instance, suppose in the above example that there is only one decision to make, but initially it is uncertain whether B, C, or B, D, or C, D will be the two agents in favor of the action while all others are against it. Then the incentives of B, C, D to form a bloc are the same: Without a bloc their probability of achieving their desired outcome is one third, with a bloc it is two thirds.

I consider both uncertainty and repeated play, so that agents with some uncertainty over preferences vote on a sequence of policy proposals. We can interpret the uncertainty about preferences in two complementary ways. First, suppose there is a time difference between the moment when agents coalesce in voting blocs, and the time of voting in the assembly. Then, when the agents make the commitment to act together they do not fully know which outcome they will prefer at the time of voting. Alternatively, in a world 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 each time.

I assume that there is some uncertainty about how agents vote, but that ex-ante it is possible to differentiate agents according to their expected preferences. The ex-ante differences in the preferences of the agents are key determinants of the strategic incentives to form voting blocs. Intuitively, agents prefer to coalesce with other like-minded voters.

Example 2 *Let there be an assembly with nine agents who must make a binary choice decision -pass or reject some policy proposal- by simple majority. Suppose that agents have uncertain preferences, so that each agent i favors the proposal with an independent probability* w_i. *Suppose* $w_k = 0.15$ *for* $k = 1, 2, 3, 4$, $w_5 = 0.5$ *and* $w_h = 0.85$ *for* $h = 6, 7, 8, 9$. *Table 1 shows the probability that the outcome coincides with the preference of a given agent, expressed as a percentage, given that the following voting blocs form: no blocs (row one); agents* $1, 2, 3$ *form a bloc (row two); agents* $1, 2, 3, 4$ *form a bloc (row three); agents* $1, 2, 3, 5$ *form a bloc (row four); and agents* $1, 2, 3, 6$ *form another bloc (row five). 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 preferences.*

Bloc	1	2	3	4	5	6	7	8	9
None	59.0	59.0	59.0	59.0	70.9	59.0	59.0	59.0	59.0
$\{1, 2, 3\}$	63.6	63.6	63.6	69.9	73.5	48.8	48.8	48.8	48.8
$\{1, 2, 3, 4\}$				67.2					
$\{1, 2, 3, 5\}$					63.7				
$\{1, 2, 3, 6\}$						43.4			

Table 1: Probability that agents get their desired outcome, in %.

The numbers on the table come from simple binomial calculations. Note that the formation of a voting bloc by agents $1, 2, 3$ has a significant effect on the outcome, even though this bloc does not command in itself a majority, unlike the bloc in the simplistic Example 1. The three agents that form the bloc increase their probability of achieving their desired policy outcome by four percentage points, so none of them have an incentive to abandon the bloc and disband it. The last three rows of the table show that no other agent has an incentive to join the voting bloc, and therefore, the formation of this bloc is a Nash equilibrium. Surely, it is not a unique Nash equilibrium: The same calculations apply if $2, 3, 4$ form a voting bloc instead, or $6, 7, 8$ among other possibilities. I propose solutions for this multiplicity below, merely noting by this example that there exist Nash equilibria in which agents form voting blocs to coordinate their votes in such a way that they affect policy outcomes to their benefit.

The insights gained in these two examples apply to voting in committees, councils, assemblies, and, in particular, in legislatures where legislators can coalesce into political parties that function as voting blocs.

Theory

Let there be an assembly N with n voters, where n is odd. The assembly chooses a policy outcome for each of T stages. Let i denote an arbitrary voter, and t an arbitrary stage. In each stage, a policy proposal is exogenously given, and the assembly makes a binary decision on whether to adopt this proposal, or reject it, in which case a default policy is implemented. Slightly abusing notation, let the policy proposal put to a vote in stage t be labeled proposal

t. At each stage, the assembly chooses the policy outcome by majority voting. The division of the assembly is the partition of the assembly into two sets: Those who vote in favor of proposal t, and those who vote against of t. The assembly makes a decision by simple majority: If the number of votes in favor is at least $\frac{n+1}{2}$, then the proposal passes, otherwise proposal t fails and the default policy is implemented at this stage. In either case, policy proposal $t+1$ is put to a vote in the following stage.

At stage t, voter i receives utility one if the policy outcome coincides with her preference in favor or against proposal t and zero otherwise. Utility is additive over time with no discounting. Since agents have neither varying intensity over preferences, nor discounting over time, their optimization problem is to maximize the number of stages in which the policy outcome coincides with their binary preference for or against the proposal. Let $p_i^t = 1$ if agent i prefers proposal t to pass, and zero otherwise; let $p^t = (p_1^t, ...p_n^t)$ indicate the profile of preferences at stage t of the whole set of voters, and let $p_{-i}^t = (p_1, ..., p_{i-1}, p_{i+1}, ..., p_n)$ be the profile without the preference of i. Similarly, let $v_i^t = 1$ if agent i votes in favor of proposal t in the division of the assembly, and $v_i^t = 0$ otherwise. Then, policy proposal t passes if and only if $\sum_{i=1}^{n} v_i^t \geq \frac{n+1}{2}$.

Agents face uncertainty at the beginning of each stage. They do not know the profile of preferences in favor or against the proposal. They only know, for any profile of preferences $p^t \in \{1,0\}^n$, the probability that p^t occurs. Let $\Omega^t : \{1,0\}^n \longrightarrow [0,1]$ be the probability distribution over preference profiles at stage t and let Ω^t be exogenous, and common knowledge at the beginning of stage t. This uncertainty is resolved and agents privately learn their own preference before the policy proposal comes to a vote. However, prior to the resolution of the uncertainty about preferences, and knowing only Ω^t, agents can coalesce into voting blocs.

Any subset of the assembly $C_l \subseteq N$ can coordinate the voting behavior of its members by forming a voting bloc $V_l = (C_l, r_l)$ with an internal voting rule r_l that maps the preferences of its members into votes cast by the bloc in the division of the assembly. I assume that joining a voting bloc is voluntary and agents may also remain independent. Formally, I assume that there exists a set of party labels $\mathcal{L} = \{0, 1, ..., L\}$, and each party label $l \in \mathcal{L}$ is associated with a voting rule r_l. Agents choose a party label. Those who choose label 0 remain independent, while for every other label $l \in \mathcal{L}$, all agents who choose the label l join V_l with internal voting rule r_l.

If a coalition C_l of size N_l forms a voting bloc with rule r_l at stage t, in an internal meeting prior to the assembly meeting, the members of C_l vote to determine their coordinated behavior in the division of the assembly according to their own internal rule r_l. In particular, each member of C_l casts an internal vote $\widehat{p}_i^t \in \{1, 0\}$ for or against the proposal, and these internal votes are aggregated into a common outcome for the bloc with three possibilities:

1. If $\sum_{i \in C_l} \widehat{p}_i^t \geq r_l N_l$, then $\sum_{i \in C_l} v_i^t = N_l$. If the fraction of C_l members who favor the policy proposal is at least r_l, then the whole bloc votes for the proposal in the division of the assembly.

2. If $\sum_{i \in C_l} \widehat{p}_i^t \leq (1 - r_l) N_l$, then $\sum_{i \in C_l} v_i^t = 0$. If the fraction of C_l members who are against the policy proposal is at least r_l, then the whole bloc votes against the proposal in the division of the assembly.

3. If $(1 - r_l) N_l < \sum_{i \in C_l} \widehat{p}_i^t < r_l N_l$, then $\sum_{i \in C_l} v_i^t = \sum_{i \in C_l} \widehat{p}_i^t$. If neither side gains a sufficient majority within the voting bloc, the bloc does not act together and it reproduces in the division of the assembly the same internal split.

I model the independent agents as if they formed a voting bloc V_0 with unanimity rule $r_0 = 1$ so that no independent ever votes against her wish in the assembly. I assume that every other voting bloc has commitment mechanisms such that the outcome of their internal meeting is binding for the vote in the division of the assembly at stage t, but the contract to join a bloc is binding for only one stage. Each agent must sign exactly one contract in each period. While the assumption of binding commitments might be restrictive in some applications, if legislators can make deposits (of money or effort) up front as guarantee that they will not defect, or if it is possible to punish defectors ex-post if only by social sanctions such as ostracizing them, then these commitment technologies suffice to sustain the coordination of votes.[3]

I assume that the set of possible internal voting rules is the set of all majority rules, from simple majority to a $\frac{n-1}{n}$ supermajority rule, and that for each rule, there exists several labels that use the same rule, so that two or more distinct voting blocs can form at the same time with an identical internal rule. Formally, for any bloc size N_l and for any natural number $x \in (\frac{N_l}{2}, N_l)$ that a bloc wishes to use as threshold to act together, I assume that there exists a label $l \in \mathcal{L}$ with rule r_l such that $r_l N_l = x$, and for any rule r_j, I assume that $\#\{l \in \mathcal{L} : r_l = r_j\} \geq \frac{n}{3}$.

The timing of stage t is as follows:

1. The probability distribution over preferences Ω^t and the set of labels $\mathcal{L}$ with their associated rules are public knowledge. Agent i remains uncertain about the exact preference p_i^t.

[3]For instance, legislators could be asked to make fundraising efforts on behalf of a party, in anticipation that all the collected resources will be shared by those who do not defect.

2. Each agent i chooses a label $l_i \in \mathcal{L}$. The set of agents C_l who choose label l form a voting bloc $V_l = (C_l, r_l)$.

3. Each agent i privately learns p_i^t, her preference for or against policy t.

4. Voting blocs meet. Each member of a bloc casts a vote for or against the proposal, and these votes, together with the internal rule of the bloc, determine the actions of the bloc in the division of the assembly.

5. The assembly meets. Agents vote according to the outcome of substage 4, and the proposal passes if it gathers a simple majority of favorable votes.

The intuition of this timing is that initially agents know who is likely to favor or oppose a bill, but they can't be sure since they have not read the details of the bill yet. At this point, agents form alliances and coalesce into groups, committing to discuss the bill internally and act as a bloc once the bill comes to the floor. Agents choose who to join based upon expected correlations in future preferences, and by choosing a particular label, a group of agents chooses to be together and at the same time they choose the rule they will use to aggregate preferences. Then a lower chamber, or a committee, or an arbitrary exogenous body, produces the bill for agents to inspect it, and voters learn their true preference. Voting blocs then meet to aggregate the preferences of their members, revealed by the votes at the internal meeting. Finally the assembly meets after blocs have committed to a coordinated voting strategy and the bill passes if it gathers enough votes.

Let h^t be the history of the game up to stage t. This history includes the probability distributions over preferences of all previous periods, $(\Omega^\tau)_{\tau=1}^{t-1}$ as well as the actions of all agents in all previous periods. Let s_i^t denote the pure stage strategy of agent i at stage t, which specifies two elements. First, given the history h^t and the current probability distribution over preference profiles Ω^t, s_i^t determines the label l_i^t that i chooses, and s_i^t also determines the vote $\widehat{p}_i^t$ of agent i for or against proposal t inside the voting bloc, as a function of the true preference of i and the labels chosen by all agents $l^t = (l_1^t, ..., l_n^t)$. Formally, $s_i^t(h^t, \Omega^t) = (l_i^t, \widehat{p}_i^t(p_i^t, l^t))$ and $h^t = (\Omega^\tau, l^\tau, \widehat{p}^\tau)_{\tau=1}^{t-1}$.

A pure strategy for the game for agent i is a finite sequence of T pure stage strategy functions, one for each stage: $s_i = (s_i^t(h^t, \Omega^t))_{t=1}^T$. A pure strategy profile for the game is then $s \in S$, where $s = (s_1, ..., s_n)$ and S is the set of all feasible strategies profiles. I denote by $s^t = (s_1^t, ..., s_n^t)$ the pure strategy profile at stage t, so note that s is also the finite sequence of stage strategy profiles $s = (s^t)_{t=1}^T$.

The stage utility of agent i, defined at the beginning of stage t, is the ex-ante probability that the policy outcome at the end of the stage coincides with the preference of the agent, given the probability distribution over preferences, and the stage strategies of every agent. I denote this utility by $u_i^t(s)$. No discounting means that the aggregate ex-ante utility of agent i, evaluated at the beginning of the first period, is equal to $\sum_{t=1}^{T} u_i^t(s)$. The goal of each agent is to maximize this aggregate utility, by choosing which voting blocs to join, and how to vote. Note that Ω^t determines the expected payoffs of the stage game t as a function of the actions of the players. Let $\Omega = (\Omega^t)_{t=1}^T$ be the sequence of probability distributions over preferences at each stage, which in turn determines the payoffs of the whole game.

Then $\Gamma = (N, S, \Omega)$ denotes the political game of coalition formation and voting that the set N of agents play in T stages.

A subgame perfect Nash equilibrium of the game Γ specifies strategies for all agents that are mutual best responses for any subgame of the game. Given the multiplicity of such equilibria with uninteresting properties (such as a trivial equilibrium where no one joins a bloc and everyone votes against every proposal, so no proposal passes and no unilateral deviation can make an agent better off), I look for equilibria that satisfy two added properties: stationarity, and weak stage-dominance.

Definition 1 *A strategy profile s is stationary if the stage strategy s_i^t is independent of history h^t for every voter i and stage t.*

The probability distribution over preferences Ω^t determines the expected payoffs of the stage game. Stationarity as defined means that faced with a given stage game at a given time, an agent uses the same stage strategy regardless of what happened at any previous stage, ruling out inter-temporal punishments of the actions taken by an agent at any previous time. In stationary equilibria, agents are able to maximize their stage utility myopically and at the same time maximize the aggregate inter-temporal utility, which reduces the complexity of the optimization problem.

Stationarity alone does not prevent implausible equilibria in which every agent votes against the proposal. In a one-shot game, such equilibria are discarded assuming that agents never play weakly dominated strategies. I extend this notion of undomination assuming that voters, while holding the strategies of all players in future stages as fixed, do not play stage strategies that are weakly dominated. At any voting stage, and considering the stage game in isolation, voting for the least preferred of the two policy alternatives is weakly dominated. *Stage undomination* (Baron and Kalai, 1993) requires that voters do not vote for their least preferred candidate at any stage unless they expect to gain something from having cast such a vote in a future stage.

Definition 2 *Let $S_s \subset S$ be the set of all strategy profiles in which exactly one insincere* *vote from strategy profile s is reversed and made sincere. The strategy profile s is stage undominated if* $\sum_{\tau=t+1}^{T} u_i^{\tau}(s) > \sum_{\tau=t+1}^{T} u_i^{\tau}(s')$ *for any strategy profile $s' \in S_s$ such that the vote reversed to sincerity is cast by agent i at stage t.*

Stage undomination together with stationarity imply sincere voting. Since there is no obvious notion of sincerity in the game of coalition formation, I say that a strategy profile is sincere if voting at the internal meetings is sincere.

Definition 3 *A strategy profile s is sincere if $\widehat{p}_i^t(p_i^t, l^t) = p_i^t$ for every voter i and stage t.*

Lemma 1 *If a strategy profile s is stationary and stage undominated, then it is sincere.*

The proof is immediate: If a strategy profile is stationary, stage strategies and stage payoffs in all future stages are, by definition, independent of the vote that an agent casts in the current stage. Since insincere voting is weakly dominated in the stage game, if the agent casts an insincere vote in the current stage, it must then do so in violation of stage undomination.

The equilibrium concept I use is Stage-Undominated Stationary Pure-Strategy Subgame-Perfect Nash Equilibrium, from now on referred simply as an *equilibrium.* While it is possible to find equilibria in which voting in the assembly is unaffected by voting blocs -either because these don't form, or if they do their members all agree in their preferences so no internal rolling takes place,- the more interesting case is the existence of equilibria in which voting blocs form and affect voting records by aggregating internal preferences in such way that exercises party discipline, so that agents cast votes in the assembly against their individual preferences.

Definition 4 *A sincere strategy profile s exhibits party discipline if it is such that, with positive probability, $v_i^t \neq p_i^t$ for some agent i and stage t.*

In a strategy profile with sincere voting, agents vote for their preferred alternative in the internal meeting of their voting bloc and, if they are independent, in the assembly. An agent who follows a sincere voting strategy only casts a vote against her preference if she loses the internal vote of her bloc.

Proposition 2 *Suppose Ω^t has full support for some t. Then there exists an equilibrium with party discipline.*

Proof. Let s be such that $s_i^t(h^t, \Omega^t) = (l_i^t, \widehat{p}_i^t(p_i^t, l^t)) = (l, p_i^t)$ with $r_l = \frac{n+1}{2n}$ for every voter i and stage t, and for any history h^t. I first show that s exhibits party discipline. Take t such that Ω^t has full support, then $p^t = (1, 0, 0, ..., 0)$ with positive probability. Agent 1 loses the internal vote, and $p_1^t = l$ but $v_1^t = 0$. Second, I note that the proposed strategy is stage undominated, since no agent ever votes internally for her least preferred alternative; stationary, since the strategy is history-independent; and pure. I now show that no deviation can make an agent i better off at any subgame. First, deviating at stage t by choosing $l_i^t \neq l$ or $\widehat{p}_i^t \neq p_i^t$ has no effect on the play or payoffs on any other stage, since s is stationary. Hence it suffices to analyze the incentives to deviate in each stage game. If agent i chooses $\widehat{p}_i^t \neq p_i^t$, either i is not pivotal and the deviation has no effect, or i is pivotal and her deviation changes the outcome and lowers her payoff. Finally, suppose that agent i deviates by choosing $l_i^t \neq l$. Note that the voting bloc acts as a dictator even after the defection of agent i. Therefore, if agent i votes against the preference of the majority of the bloc, agent i loses and the outcome is the same as if i stays and loses inside the bloc; if agent i leaves the bloc and votes with the preferences of at least one half of the remaining members of the bloc, then she wins, but she also wins if she stays in the bloc, so she gains nothing by deviating. In either case, agent i has no incentives to deviate at stage t, for any t or any combination of stages. ■

In the following section of the paper I sharpen this existence result in a stylized assembly. First I present a second general result, which is important to interpret the endogenous voting blocs that emerge in each stage as permanent, stable political parties. Let Γ^1 denote the one-shot game with preferences given by Ω^1. This is stage game $t = 1$ considered in isolation, disregarding future stages. In a repeated finite game, the strategy consisting of playing a Nash equilibrium of the stage game at each stage is a Nash equilibrium of the repeated game. This is a fairly well-known result in repeated game theory (Fudenberg and Tirole, 1991, p 149). Applying this argument to my model, I find that maintaining the same voting blocs over time is an equilibrium strategy as long as the preferences that sustain such voting blocs do not change.

Proposition 3 *Suppose $\Omega^t = \Omega^1$ for any $t = 1, 2, ...T$. Let the strategy profile s^1 be an equilibrium of the stage game Γ^1. Then the strategy profile s consisting of playing s^1 at any stage after any history is an equilibrium of the repeated game with T stages.*

As long as the priors over preferences of the members of the assembly given by Ω^1 do not change, it remains an equilibrium of the stage game for the same voting blocs to persist over time. I interpret these long-lasting, permanent voting blocs that exercise party discipline as political parties, which only break up when a change in preferences makes the current division into parties unstable. Note that this permanence of the same voting blocs in equilibrium is not trivially implied by stationarity. Stationarity requires that at a given stage, the stage formation of blocs must be the same for any history as long as the probability distribution over preferences is the same, but it does not rule out a different configuration of voting blocs at each stage.

Endogenous Voting Blocs in a Small Assembly

Consider an assembly with nine voters, so the policy proposal passes if it gathers five favorable votes. Suppose that agents have independent preferences. That is, Ω^t is uncorrelated, each agent i has a prior w_i^t, which is the probability that i favors proposal t, and the preference profile p^t occurs with probability $\prod_{i=1}^{9} w_i^t p_i^t + (1-w_i^t)(1-p_i^t)$. Equivalently, $\Pr[p_i^t = 1 | p_{-i}^t] = w_i^t$ for all i, t and p_{-i}^t.

Suppose further that Ω^t is fixed over time and the distribution of priors is symmetric as follows: w_1=w_2=0.5-α-β; w_3=w_4=0.5-α; w_5=0.5; w_6=w_7=0.5+α; w_8=w_9=0.5+α+β, with $\alpha, \beta \geq 0$ and $\alpha + \beta \leq 0.5$. In words, priors are symmetrically distributed around one half.

The parameters α and β have an intuitive interpretation: α measures the polarization of preferences within the assembly. A hypothetical coalition of moderates comprising agents 3 through 7 (enough to become a majority centered around the median) spans an interval of priors of length 2α. A more polarized assembly corresponds to a larger α and larger differences in priors that a coalition of moderates must accommodate in order to form a voting bloc. The parameter β, albeit crudely, reflects the heterogeneity in types within each side of the assembly, or in other words, the extremism of the left-most and right-most wings.

An intuitive conjecture is that intense polarization in the assembly would make a central voting bloc unstable and would induce the formation of two opposing voting blocs, one on each side of the median. To formulate a theoretical prediction using the model, given that Ω^t is fixed over time, I take advantage of Proposition 3 to look only at the equilibrium of a simplified one-stage game, constructing the equilibrium of the whole game as the repetition of the stage equilibrium at all stages. I can then drop the superindex t from all variables without confusion.

Proposition 2 informs us that equilibria with party discipline exist. There often exists multiple equilibria, but some are more interesting than others. I seek equilibria such that party discipline not only affects voting records by rolling a few isolated, non-pivotal votes, but rather, the formation of voting blocs affects policy outcomes, by making the majority of votes in the division of the assembly not correspond with the majority of sincere preferences. In these equilibria, party discipline is relevant for the policy outcome.

Definition 5 *A sincere strategy profile s exhibits relevant party discipline if either* $\left(\sum_{i=1}^{n} p_i < \frac{n+1}{2} \leq \sum_{i=1}^{n} v_i\right)$ *or* $\left(\sum_{i=1}^{n} v_i < \frac{n+1}{2} \leq \sum_{i=1}^{n} p_i\right)$ *occurs with positive probability.*

Example 2 considered an assembly with $\alpha = 0.35$ and $\beta = 0$. Note that there are several equilibria with relevant party discipline in this assembly: Agents $1, 2, 3$ may form a voting bloc, or agents $2, 3, 4$, or instead agents $6, 7, 8$ or $7, 8, 9$ can form a unique voting bloc in equilibrium. If any of these blocs form, no agent has an individual incentive to deviate. However, all these equilibria are intuitively unsatisfactory. If the agents with a low prior form a voting bloc to their advantage and to the detriment of the agents with a high prior, it begs the question: why don't agents with a high prior form their own voting bloc as well? The game-theorist response is that under Nash equilibria, agents can only deviate unilaterally, taking the strategies of other agents as given. Table 2 adds one more row to

Table 1, to compare the outcome if both the agents with a low prior and the agents with a high prior form a voting bloc.

Blocs	1	2	3	4	5	6	7	8	9
{123}	63.6	63.6	63.6	69.9	73.5	48.8	48.8	48.8	48.8
{123}, {789}							53.3	53.3	53.3

Table 2: Table 1 extended to consider two voting blocs.

If three high-prior agents form a voting bloc in response to the low-prior voting bloc, they increase the probability that the policy outcome coincides with their individual preference from a bit less than one-half, to a bit more than one-half. But under Nash equilibria, agents best-respond unilaterally, and agents $7, 8, 9$ cannot coordinate a coalitional deviation away from the equilibrium with only one bloc, even though the outcome with two blocs is itself an equilibrium. Hard as it may be for agents to communicate and coordinate across preexisting blocs, it seems easier to scheme deviations involving only independents or agents of a single bloc, or a mix of both defecting together and possibly forming a new voting bloc. The following equilibrium concept allows for a coalitional deviation in which one bloc faces a split, a number (possibly zero) of its members defect, and at the same time a (possibly empty) subset of the defectors and previously independent agents form a new voting bloc.

Definition 6 *An equilibrium strategy profile s is split-proof if there exist no set of agents E, label $l \in \mathcal{L}$ and strategy profile $s' \in S$ such that:*

(i) $l_i \in \{0, l\}$ for all $i \in E$,

(ii) $l'_i \in \{0, l'\}$ for all $i \in E$, where $l' \in \mathcal{L}$ is such that for any voter $h \in N$, $l_h \neq l'$.

(iii) $s_h = s'_h$ for any $h \notin E$, and

(iv) for any $i \in E$ s.t. $l'_i = 0$, $u_i(s') \geq u_i(s)$ and for any i s.t. $l'_i = l'$, $u_i(s') > u_i(s)$.

Condition *(i)* says that all the agents who coordinate a deviation are initially either members of the bloc V_l, or independents. Condition *(ii)* states that after the deviation, all the deviants become either independents, or members of a new bloc $V_{l'}$ that previously had attracted no membership. Condition *(iii)* states that the rest of the agents do not react to the deviation in any way. Condition *(iv)* states that agents who defect become better off. When agents are indifferent between deviating or not, this fourth condition incorporates an intuitive discrimination: Agents may abandon a bloc to become independents when indifferent, but they only deviate to a new bloc for a strict improvement. That is, agents break indifference as if they had a lexicographic preference for independence.

The intuition for the split-proof equilibrium is that coalitional deviations across blocs are harder to coordinate, perhaps because communication is limited across blocs, or because different blocs are antagonistic and suspicious of each other, while a disaffected subset of a bloc can more easily break apart and possibly recruit some independent agents for a new voting bloc. Note that under the split-proof refinement, a whole bloc can also deviate to change their internal rule and choose a new one, therefore in a split-proof equilibrium no bloc uses a given rule when all its members would prefer to use a different one.

Related concepts of equilibrium are the Coalition-Proof equilibrium (Bernheim, Peleg and Whinston, 1987) and the Equilibrium Binding Agreements (Ray and Vohra, 1997). A split-proof equilibrium is different from these because it is not a recursive concept and it allows deviants to coordinate with independents. Under split-proof equilibria, agents negotiate as if each coalition was in its own room, but the independents were all in a central lobby, so that when a set of deviants departs from a coalition they can recruit any number of independents in their way to a new room. The split-proof equilibrium follows more closely the *split stability* notion (Kaminski, 2001; Kaminski, 2006). Parties satisfy *split stability* if they have no incentives to dissolve into smaller units, where the incentives are considered non-recursively. The split equilibrium in this paper requires not only that members of a party have no incentives to leave the party, but also that no subset of independents has an incentive to deviate and form a new voting bloc, either by themselves, or attracting defectors from one existing party.

In this section I find split-proof equilibria with connected voting blocs and relevant party discipline.

Definition 7 *A voting bloc V_l is **connected** with respect to the order $<$ if for all voters h, i, k, $(l_h = l_k = l$ and $h < i < k)$ implies $l_i = l$.*

Intuitively, the voting bloc V_l is connected if given any pair of agents who choose label l, any other intermediate agent located between the original pair also chooses label l. The order I use to define connectedness is according to the priors over preferences, w_i. A voting bloc is connected if its members are in consecutive positions in the ordering by priors. See (Axelrod, 1970) for a detailed argument in favor of connected coalitions over non-connected ones.

Using numerical simulation for a fine grid of values of α and β, I find the solutions to the game. I assume that the list of labels available to the agents includes label 0 for independents, and at least three different labels for each of the following rules: $\frac{2}{3}, \frac{3}{4}, \frac{3}{5}, \frac{4}{5}, \frac{5}{6}, \frac{4}{7}, \frac{5}{7}, \frac{6}{7}, \frac{5}{8}, \frac{7}{8}, \frac{5}{9}, \frac{7}{9}$ and $\frac{8}{9}$, so that agents can form blocs of any size with any desired supermajority rule. The rule $\frac{5}{9}$ corresponds to simple majority for any bloc of any size.

The intuition that in a very polarized assembly there is no unique moderate bloc, but rather, two blocs one in each side of the political spectrum is verified. I abuse notation to let L denote the "Left" party.

Result 4 *There are only four split-proof equilibria with connected voting blocs and relevant party discipline. In each equilibrium, exactly two blocs* $V_L = (C_L, r_L)$ *and* $V_R = (C_R, r_R)$ *form, both with three members and simple majority internal voting rule, and* $C_L \subset \{1, 2, 3, 4\}$, $C_R \subset \{6, 7, 8, 9\}$.

There are no split-proof equilibria with relevant party discipline in which a unique voting bloc forms, because if a unique bloc is of majority size or is small but moderate, other agents deviate to join it, while if the bloc is small and extreme, agents on the opposite side deviate to form a second bloc. In equilibrium, three of the four members of the assembly with a low prior choose to form a voting bloc with simple majority, and three members with a high type form another bloc. It is easy to visualize the V_L bloc as a "left" or pro-status quo party, which tends to vote against the policy proposal, and the V_R bloc as a "right" or reform party, which tends to vote for the policy proposal.

[FIGURE 1 ABOUT HERE]

Figure 1a,b,c shows in black the parameter values for which each of the four outcomes is a split-proof equilibrium. For any $\alpha < 0.5$, the voting blocs exercise relevant party discipline.

The three figures share the common characteristic that the equilibria with relevant party discipline holds only for a high α. If the assembly is not polarized and agents share similar priors, then each agent in voting bloc V_L has an incentive to defect to V_R, effectively disbanding V_L since no bloc can function with only two members. If there is enough polarization, defections across blocs no longer occur.

If the assembly is sufficiently polarized, there is a split-proof equilibrium with connected voting blocs and relevant party discipline in which two opposing blocs form, one at each side of the median.

Robustness Check

In the rest of the section I depart from the stylized assumptions of the modelled assembly, to check that the theory also predicts that two opposing blocs form in an assembly where agents do not have symmetric and independent types, but rather, their preferences are correlated and the distribution of these preferences in the assembly is asymmetric.

I look at real data from the United States Supreme Court and I calculate the effect of hypothetical voting blocs upon the outcome of the Court. This empirical exercise does not test the theory. I treat the recorded votes as if they revealed the true preferences of the justices and I use the theory to calculate what parties would have formed in the Court in a counterfactual in which justices had adopted partisan strategies.

For a more thorough discussion of the Supreme Court and the data analysis on the counterfactual effect of blocs in the Court, please consult the online Supporting Information section. Here I merely summarize the findings.

I use the data on the 419 non-unanimous decisions of the full Court from 1995 to 2004 from *The United States Supreme Court Judicial Database* (Spaeth, 2006), where the votes of each justice are coded as zero or one depending on whether the vote to affirm or reverse the decision of the previous court is interpreted as more liberal or more conservative. The nine justices, ordered from least to most conservative, are: Stevens, Ginsburg, Souter, Breyer, O'Connor, Kennedy, Rehnquist, Scalia and Thomas.

I calculate the changes on votes and outcomes that would have been brought by any hypothetical set of voting blocs that are connected according to this order

To calculate the effect of voting blocs upon the utility of the justices, it is necessary to make an assumption about the utility function of the justices. I assume that justices are outcome oriented: Each individual justice has policy preferences over the outcome of each decision and wants the Court to reach a decision according to the preference of the justice. Abstracting from the rich nuances of written opinions, the part of the outcome in each case that is apt for quantitative analysis is the binary decision to either affirm or reverse the ruling from a lower court, to side either with the plaintiff or the defendant. I assume that each justice prefers one of these two outcomes over the other and each justice gets a higher utility if his preferred outcome is the one selected by the Court by majority voting. Then I assume that for the aggregate of all 419 cases from 1995 to 2004 the goal of each justice was to maximize the number of cases in which the decision of the Court coincides with the preference of the justice, as revealed by the vote of the justice in the data.

For any given configuration into voting blocs in the Court, imagine that each bloc holds a private internal vote before the division of the Court, and in these internal votes I imagine that each justice votes according to how the justice voted in reality in that case. Then I aggregate the votes inside each bloc according to the majority rule of the bloc, and I calculate the new outcome in the division of the Court, once I take into account that some justices now cast a vote against their preference along the lines dictated by the majority of their bloc. Finally, I calculate how many decisions change with the voting blocs under consideration relative to the original data, and for each justice I calculate the net balance of decisions that change to favor her preferences minus the number of decisions that change against her preference. In this manner I assign a vector of payoffs for each possible configuration of voting blocs, and then I can check which of these configurations are equilibrium predictions if justices formed blocs strategically.

Result 5 *There exist split-proof equilibria such that any three of the four most liberal justices (Stevens, Ginsburg, Souter, Breyer) form a voting bloc and Rehnquist, Scalia and Thomas form a second voting bloc. There is no other split-proof equilibrium with connected voting blocs.*

According to Result 5, the equilibrium outcomes are such that two opposing blocs -one at each side of the ideological spectrum- counterbalance each other, and the swing moderate agents, in this case O'Connor and Kennedy remain unaffiliated, independent. Compare Result 5 with Result 4 in the stylized assembly. The equilibrium predictions with the US Supreme Court data are a subset of the predictions for the idealized assembly. The theoretical model predicted two voting blocs of size three, one with three of the four most liberal members, the other with three of the four most conservative members. The prediction with the empirical data fits within the stylized prediction, and the only difference is that the conservative bloc has to be $\{789\}$ and not $\{678\}$.

Result 5 shows what voting blocs would occur in equilibrium in an assembly with nine rational agents who are strategic and can coordinate their votes without constraints, and whose preferences are consistent with those revealed by the pattern of votes in the US Supreme Court for 1995-2004. It has not provided, nor did it intend to provide, a theory of voting in the Court. Members of a committee or assembly with size and preferences identical to those of the US Supreme Court face strategic incentives to coalesce into voting blocs. An explanation of whether or not the US Supreme Court justices act upon these strategic incentives is beyond the scope of this paper.

Conclusion

Members of a democratic assembly -legislature, council, committee- can affect the policy outcome by forming voting blocs. A voting bloc coordinates the voting behavior of its members according to an internal voting rule independent of the rule of the assembly, and this coordination of votes affects the outcome in the division of the assembly.

I have shown that for any assembly there exists an equilibrium in which agents endogenously form voting blocs and these blocs affect voting records exercising party discipline. I have further shown that permanent voting blocs -parties- persist over time in equilibrium until there is a change in the underlying heterogeneous preferences.

In a small assembly, I predict that two polarized voting blocs form. I first derive this prediction from an abstract assembly with stylized preferences, but I obtain the same prediction when I assign preferences according to the empirical data on voting records in the United Supreme Court between 1995 and 2004. In either case, the outcome of a coalition formation game in which agents strategically formed voting blocs would result in two voting blocs, one at each side of the median member of the assembly.

I have shown that strategic members of an assembly form political parties to coordinate their votes. I explain political parties as endogenous voting blocs in an assembly. This motivation to form parties is complementary to the traditional views of parties as electoral machines, but it is simpler, because it does not rely on outside agents like citizens or lobbies, and it is also more general, since it can also explain the formation of parties in unelected assemblies, such as the House of Lords in the United Kingdom, the formation of blocs in the assemblies of international organizations such as the UN or the IMF, or the formation of European parties in the European Parliament, parties that vote together but nevertheless fracture into independent national parties for electoral purposes.

References

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

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

Ashworth, Scott and Ethan Bueno de Mesquita. 2008. "Informative Party Labels With Institutional and Electoral Variation." *Journal of Theoretical Politics* 20(3):251.

Axelrod, Robert M. 1970. *Conflict of Interest: A Theory of Divergent Goals with Application to Politics.* Chicago, Markham.

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

Baron, David and Ehud Kalai. 1993. "The Simplest Equilibrium of a Majority Rule Division Game." *Journal of Economic Theory* 61:290–301.

Bernheim, B. Douglas, Bezalel Peleg and Michael D Whinston. 1987. "Coalition-Proof Nash Equilibria: I. Concepts." *Journal of Economic Theory* 42(1):1–12.

Carey, John M. 2007. "Competing Principals, Political Institutions, and Party Unity in Legislative Voting." *American Journal of Political Science* 51(1):92–107.

Carruba, Clifford J and Craig Volden. 2000. "Coalitional Politics and Logrolling in Legislative Institutions." *American Journal of Political Science* 44(2):261–277.

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

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

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

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

Eguia, Jon X. 2010. "Voting Blocs, Party Discipline and Party Formation." SSRN wp 1514506.

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

Fudenberg, Drew and Jean Tirole. 1991. *Game Theory.* Cambridge, MIT Press.

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

Kaminski, Marek M. 2001. "Coalitional Stability of Multi-Party Systems: Evidence from Poland." *American Journal of Political Science* 45:294–312.

Kaminski, Marek M. 2006. "A General Equilibrium Model of Multi-Party Competition." *Social Choice and Welfare* 26:333–361.

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

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

Ray, Debraj and Rajiv Vohra. 1997. "Equilibrium Binding Agreements." *The Journal of Economic Theory* 73:30–78.

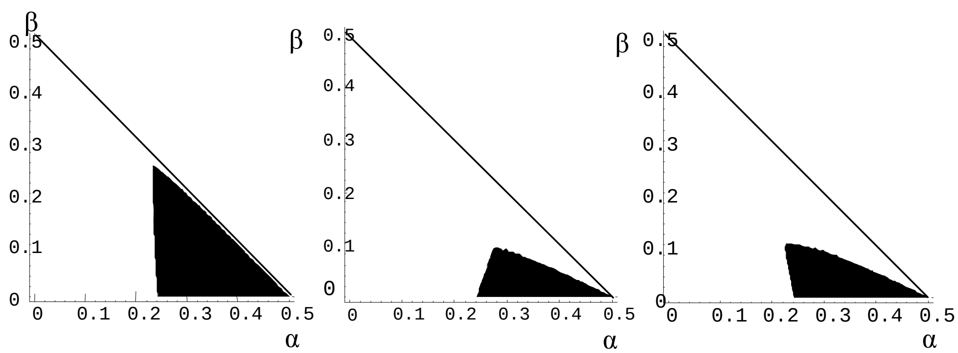
Rohde, David W. 1991. *Parties and Leaders in the Postreform House.* Unversity of Chicago Press.

Schattschneider, Elmer E. 1942. *Party Government.* Rinehart.

Schwartz, T. 1989. "Why Parties?" *Research memorandum, Dept. Political Science, UC Los Angeles* .

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

Spaeth, Harold. 2006. “United States Supreme Court Judicial Database 1953-2005.” Computer File available at www.as.uky.edu/polisci/ulmerproject.



$C_L=\{2,3,4\}$, $C_R=\{6,7,8\}$.

$C_L=\{1,2,3\}$, $C_R=\{6,7,8\}$ or $C_L=\{2,3,4\}$, $C_R=\{7,8,9\}$

$C_L=\{1,2,3\}$, $C_R=\{7,8,9\}$

Figure 1: a,b,c. Split-proof equilibrium voting blocs.