A Theory of Elections and Voting Blocs

Thesis by

Jon X. Eguia

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2007

(Defended May 2, 2007)

Acknowledgements

I earnestly thank Tom Palfrey and Matt Jackson for many years of excellent advice, encouragement, and continuous support. I am also very grateful to Andrea Mattozzi, Chris Chambers, Federico Echenique, Matias Iaryczower, Rod Kiewiet, and Kim Border for their helpful comments and suggestions. Furthermore, I have greatly benefited from the generous contribution of time, effort, ideas, and wisdom made by many of my fellow graduate students, in particular PJ Healy.

I also wish to thank the Division of Humanities and Social Sciences and the Social and Information Sciences Laboratory at Caltech, the Olin Foundation, the Haynes Foundation and Peter Bossaerts for financial support, and the Laboratory for Experimental Social Sciences and the Department of Economics at Princeton University for their hospitality during my visit in 2005.

Abstract

I study how a group of agents with incomplete information about their conflicting preferences make a collective decision by means of voting.

I present a model of representative democracy with citizen candidates in which the set of agents who runs for office is endogenously determined. I show that if the electorate is large enough and agents are not able to perfectly anticipate the electoral outcome, elections are always contested, and an equilibrium with two candidates exists.

In the last two chapters of this dissertation, I introduce a model of voting bloc formation in which groups of agents choose to coalesce to vote together in an assembly. Looking first at one coalition, then at two coalitions, and finally at an arbitrary number of coalitions emerging in a fully endogenous model of voting bloc formation, I analyze the incentives to join a voting bloc, the stability of different voting blocs and how the incentives and stable outcomes change with the size of the blocs, the internal voting rule that each bloc uses and the heterogeneity in the preferences of the agents. This model provides a new explanation of the formation of political parties in legislatures.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A rational agent making an individual decision chooses her preferred outcome among all the feasible alternatives. A group of agents making a collective decision faces a more complicated problem, because the group preference may be difficult to determine if there is a conflic of individual preferences inside the group.

Social choice theory studies how a group of agents aggregate preferences to make a collective choice. In a democracy, the aggregation of preferences to make a collective decision is often made by means of voting. In a direct democracy, the members of the group vote over the different alternatives and a constitutionally given voting rule aggregates the votes to choose an outcome. In an indirect democracy, agents do not vote directly over the alternatives. Rather, agents vote to elect a representative, who is then endowed with the powers to make a decision on behalf of the whole group. Either directly or indirectly, agents express their preference through their votes, and a voting rule translates the votes into a collective choice. For instance, in referenda citizens of a society directly choose policy outcomes, while in regular elections they first elect politicians who then represent the citizens and choose policy.

I study the aggregation of preferences by a group of agents who face some uncertainty about the distribution of preferences inside the group and must make a collective decision to choose a policy outcome. I study the aggregation of preferences either indirectly by means of elections, or directly by means of voting over the set of alternatives.

In chapter 2 I study a representative democracy where citizens determine policy outcomes indirectly by first electing a representative. Every member of the society chooses whether or not to become a candidate -to run for election- and once the set of candidates is determined, every member of the society casts a vote for one of the self-declared candidates. The candidate who receives the most votes wins the election and chooses the policy outcome for the society. This modeling approach in which every citizen makes a strategic decision on whether or not to become a candidate is part of the "citizen candidate" literature of representative democracy sparked by the seminal work by Besley and Coate [9] and Osborne and Slivinski [44]. I improve on previous results by incorporating uncertainty into the theory. Existing models assumed that in equilibrium every agent is perfectly able to anticipate the result of the election; consequently, there could not be a two-candidate election in which one of the candidate loses because if so, the would-be loser, anticipating defeat, would choose not to run for election given that running is costly. Once we recognizing that there is an element of uncertainty in every election and that no candidate is able to perfectly predict the exact tally of votes, the theory yields better predictions. In particular, I show that if the number of voters is large enough and there is an intrinsic benefit of being elected, there are no uncontested elections and at least two candidates run for office. Furthermore, equilibria with exactly two candidates always exist. To sharpen this prediction, I introduce a refined equilibrium concept that relaxes the assumption that all agents make a strategic decision when they cast their vote, considering instead that some agents may vote sincerely for their most preferred candidate. Selecting equilibria that are robust regardless of whether any set of agents behaves sincerely or strategically at the voting stage, I find that if the policy space is one dimensional, in a two-candidate election both candidates must be moderates with one of them on each side of the ideological left-to-right spectrum.

I study the aggregation of preferences by means of voting in a direct democracy in chapter 3 and chapter 4. Given a set of alternatives and a voting rule, every agent in the society casts a vote for one of the policy alternatives. The aggregation of these votes according to the voting rule of the society determines the collective choice of a particular policy for the society. For instance, two salient rules are simple majority and unanimity. Under simple majority, the society adopts a given policy if more than a half of the agents vote in favor of it, otherwise a status quo or default policy remains in place. Under unanimity, the society adopts a given policy only if every vote favors it, otherwise, if at least one agent votes against it, the policy proposal is rejected and the status quo remains in place. Agents have incentives to behave strategically regardless of the voting rule in place, coordinating their actions with those of other agents. While the coordination of voting strategies remains difficult in a large election, the model applies to voting in committees or assemblies where agents know each other, communicate with each other and have an opportunity to coordinate with each other forming coalitions and dividing the assembly into cohesive subsets.

In legislative bodies such as lower or upper houses of representatives, senates, parliaments or national assemblies, legislators typically coalesce into cohesive political parties that discipline the voting behavior of their members. Other applications range from the formation of alliances of countries in international voting bodies, such as the European Union or the Arab League in the UN General Assembly, to the formation of factions that vote together in faculty meetings in an academic department. In all these applications, individual agents are able to communicate and coordinate with others. Specifically, agents can coalesce to form voting blocs.

A voting bloc is a set of agents who commit to vote together, aggregating their internal preferences into a common policy position that the whole bloc votes for, rolling internal dissent according to a voting rule internal to the bloc. I study the strategic incentives and the advantages of forming a voting bloc.

In chapter 3 I consider an exogenously given coalition of agents who can form a voting bloc if every coalition member agrees to the formation of the bloc. Assuming that the goal of each agent is to maximize the probability that the final policy outcome coincides with the preference of the agent, forming a voting bloc generates a net aggregate surplus of utility for the members of the coalition. However, since every member must agree in order for the bloc to form, it must be that every member individually benefits. I show under which conditions the coalition can form a voting bloc that benefits every member. I show how the selection of the voting rule internal to the bloc and used to determine the bloc's vote in the assembly given the preferences of its members affects the incentives of each agent to join the bloc or not. Specifically, I show that under some parameters, the voting bloc must choose a supermajority internal voting rule and not simple majority in order to benefit every one of its members. I find that even if every member of the coalition must agree in order to form a voting bloc -that is, each member of the coalition has veto powers over the constitutional decision to form a voting bloc- the formation of a bloc hinges on a single well-defined agent who is the least likely to benefit from the aggregation of votes in the bloc. With this result in mind, I investigate the possibility of allowing this agent to opt out of the voting bloc: Suppose that every member of the coalition must agree to the formation of a voting bloc that excludes the one agent least likely to benefit from joining the bloc. Note that the excluded agent must also agree to the formation of the bloc, in the same spirit as the United Kingdom or Denmark agreed, by the Treaty of Maastricht (1992), to the formation of the European Monetary Union that excluded them and brought the euro into existence. I find that in some circumstances, forming a voting bloc excluding one agent is beneficial for every member of the coalition, including the excluded one, even in some cases when the coalition cannot reach consensus to form a voting bloc with all its members. Overall, chapter 3 shows that forming a voting bloc results in a net gain for the members of a coalition who agree to form it, and studies some possible ways to select the internal voting rule to aggregate preferences within the bloc in such a way that every agent in the coalition participates in the gains brought by the voting bloc.

In chapter 4 I extend and generalize the theory to properly consider a fully endogenous model of voting bloc formation. Chapter 3 detailed the strategic incentives to form a voting bloc in an exogenously given coalition, given the behavior of all other agents in the assembly. This is a partial equilibrium analysis that neglects the possibility that other groups of agents outside the coalition under consideration can also form their own voting blocs. The generalized model in chapter 4 extends the analysis to study the strategic interactions between multiple voting blocs that compete with each other in their quest to influence the collectively chosen policy outcome. This model provides a novel interpretation of the process of the formation of political parties inside a legislature where agents coalesce solely for ideological reasons to attain the policy outcomes they prefer in an ideological space, absent any distributive or electoral component in their strategies. As such, it is a new insight into the formation of political parties.

The first step toward fully endogenizing the model of voting bloc formation is to consider that there exists two political parties in the assembly, each of which can form a voting bloc. I consider the case where consensus is necessary to form a voting bloc, and the case where the constitutional rules of the party are such that a subset of the party can form its own voting bloc over the opposition of some members who cannot veto the project. In this latter case I look for conditions under which every member would choose to join the voting bloc and there would be no party member who chose to be left out. I show that many results in chapter 3 generalize to the case with two competing parties: Forming a voting bloc generates a benefit and under some parameters the voting bloc must choose a supermajority internal rule to attract every member to the bloc. While in chapter 3 I note that allowing one agent to opt-out and be excluded sometimes makes the formation of a bloc possible, now I find the converse result to be also true: Under different parameters, agents would individually choose to stay out of the bloc, but if the bloc changes its formation rules so that the participation of every party member is required in order to form the bloc, every party member would join and the bloc would form. Considering both results together I conclude that while allowing a reluctant agent to stay out of the bloc may in some cases benefit everybody, allowing agents to freely enter or exit may often result in free riding and the loss of the aggregate gain that the party members derive from forming a bloc.

The culmination of the theory in this thesis provides a fully endogenous model of party formation. Agents are able to coalesce freely in the assembly, and the resulting coalitions partition the assembly into voting blocs. In an application to a legislative assembly, political parties are thus explained as the stable partition of the legislative assembly into voting blocs. This partition, along with the internal voting rules of each agent, determines the voting bloc structure of the assembly. I show that stable voting bloc structures exist for several definitions of stability, including Nash stability, where each agent is free to exit the bloc she belongs to and join any other bloc at wish. I show that Nash stable voting blocs must be of size less than minimal winning. To illustrate the model at work, and provide sharper predictions, I look at a small assembly, of fixed size nine, and I introduce a new stability concept that captures coalitional deviations in which at most one bloc splits and at most one new bloc forms. I call this notion "Split stable." I show that the only Nash and Split stable voting bloc structures in a stylized small assembly consist of two small voting blocs, one at each side of the ideological spectrum, while the median member of the assembly remains independent. I also look at voting data from the United States Supreme Court and, taking the voting records from 1995 to 2004 as indicative of the true preferences of the justices, I conclude that in an assembly with such preferences in which agents strategically formed voting blocs with the purpose of maximizing the probability that their preferred policy is implemented, the only stable voting bloc structure that results would be for two blocs to form, one composed of any three of the four liberal justices, the other one with the three most conservative justices.

The model of voting bloc formation provides an explanation for the formation of political parties, a prediction about the parties that would form in a small assembly, and a method for evaluating the merits of different rules for a coalition of agents who must select an internal rule to aggregate their preferences before they commit to vote together in a bloc.

Chapter 2

Citizen Candidates under Uncertainty

ABSTRACT: In this chapter I make two contributions to the growing literature on "citizen candidate" models of representative democracy. First, I add uncertainty about the total vote count. I show that in a society with a large electorate, where the outcome of the election is uncertain and where winning candidates receive a large reward from holding office, there will be a two-candidate equilibrium and no equilibria with a single candidate. Second, I introduce a new concept of equilibrium, which I term "sincere-strategic," and I show that with this refinement, the two equilibrium candidates will not be too extreme, one will lean to the left and the other one to the right.[1]

2.1 Introduction

In a representative democracy, citizens elect representatives who in turn, choose policies for the society. Traditional models divide the members of the society into two classes: voters, whose only political role is to vote, and politicians or political parties (often just two of them), who compete in the election. Citizen candidate models of the electoral process, on the other hand, explain how politicians emerge from the class of voters. In these models, some citizens become politicians by

[1] Part of this chapter appears as an independent paper in Eguia [22].

choosing to run as candidates in an election. The number and policy preferences of the candidates who run in equilibrium are determined by three factors: the policy preferences of every citizen, the benefits of holding office, and the cost of running for election.

The standard citizen candidate models suffer from a simplistic assumption that leads to unrealistic predictions. The assumption is that candidates can perfectly anticipate the outcome of the election and forecast exactly how many votes each candidate will receive. This assumption leads to the prediction that two candidates will run against each other only if they have the exact same number of supporters in the electorate. In models with a finite electorate, this implies that typically there will be a two-candidate equilibrium only if the number of citizens is even, because if it is odd, one of the candidates will generally have at least one more supporter than the other.

I solve this problem and assure political competition in the model by introducing uncertainty in the electoral outcome, and then showing that in a society with a large electorate, equilibria with one candidate do not exist, while equilibria with two candidates always exist. I model the uncertainty as in Myerson [43]: Citizens choose which candidate to support, but each citizen has a small probability of failing to convert her intention to support a candidate into an actual valid, counted vote for the candidate -perhaps the voter is unable to make it to the polling station, or she misuses the voting equipment and casts an invalid ballot. This individual probability of being unable to cast a valid vote generates an aggregate uncertainty about the total vote count.

The aggregate uncertainty about the electoral outcome crucially affects elections with a large electorate. Candidates decide to run based on the support they have in the electorate at the time they make the decision to run. However, elections are not deterministic: the aggregate uncertainty makes the outcome stochastic, and a candidate who initially had less support may ultimately collect more votes and win the election. If the most popular candidate enters the race, another citizen with slightly less support will also run, in the hope of an upset victory. As a result, in a society with a large enough electorate, one-candidate equilibria do not exist, whereas two-candidate equilibria always exist, regardless of the exact number of citizens. These predictions are similar to those in

Osborne and Slivinski's [44] model of citizen candidates, but Osborne and Slivinski work only with an infinite number of sincere voters, who have rather restricted preferences. My model, by contrast, allows for any finite number of citizens, who vote strategically and may have quite general forms of preferences.

The uncertainty I introduce in my model has a large effect unless the electorate is very small. In the case of a very small electorate the predictions of my model are similar to those of Besley and Coate [9]: The existence of equilibria with one or two candidates depends on whether the number of citizens is odd or even.

There are two other papers incorporating uncertainty to a model of citizen candidates: In Riviere's paper [50], a group of citizens learn their policy preference only after candidacies are announced, so the location of the median is uncertain at the time of the announcement. My model differs from Riviere's both in the assumptions I use and the results I obtain. In words of the author, Riviere's assumptions are "very restrictive" and in her citizen candidate model, two-candidate equilibria are "very rare."[2] I relax and generalize most of Riviere's assumptions to show that under mild conditions, equilibria with two candidates exist.

In a short note, Roemer [51] tackles the problem of indeterminacy of equilibria in Besley and Coate's model, advocating a particular refinement (Party-Unanimity Nash Equilibrium or PUNE) that yields a smaller but non empty set of equilibria. Uncertainty is only a side issue in this note, and Roemer assumes that each candidate wins with a probability equal to the candidate's share of votes. His main conclusion is that "the citizen candidate equilibrium concept is not sufficiently restrictive."

I address the multiplicity of equilibria by proposing a more restrictive equilibrium concept. This refined equilibrium, which I term *sincere-strategic*, relaxes the assumption that citizens are strategic in their voting behavior, allowing for citizens to choose which candidate to support either sincerely

[2] Riviere also considers a game in which like-minded citizens can share the cost of running, forming a political party. In this setting, equilibria with two parties are no longer rare, but for some values of the cost of running, only equilibria with one or three candidates exist.

or strategically. I show that under this sincere-strategic equilibrium concept, all candidates must receive a similar share of electoral support and that if there are only two, one will lean left and the other one will lean right, but they will not be too extreme.

The rest of the chapter is as follows: In section 2.2 I present the model of citizen candidates with uncertainty. In section 2.3 I characterize equilibria with one and two candidates, in pure and mixed strategies. In section 2.4 I introduce the refined sincere-strategic equilibrium, and I apply this concept to equilibria with two candidates and multiple candidates. I summarize the findings of the chapter in section 2.5. An appendix contains technical matters omitted from the text.

2.2 The Model

Let $\mathcal{N}$ be a society formed by N citizens labeled $i \in \mathcal{N} = \{1, ..., N\}$, with $N \geq 3$. This society must elect a representative, who will receive a benefit b from being elected and will also get to choose a policy in a unidimensional policy space $[0, 1]$. The implemented policy I denote by p. Citizens have different preferences over the chosen policy; let $v_i(p)$ be a bounded function measuring the utility that citizen i derives from policy $p \in [0, 1]$.

Each agent in the society can run as a candidate in the election, but doing so entails a cost c, which is small compared to the rewards from holding office. Formally, I assume that $b > 2c > 0$.

Let $I_i = 1$ if i runs as a candidate and zero otherwise, and let $W_i = 1$ if i runs as a candidate and wins the election and zero otherwise.

Then the utility of agent i is:

$$U(p, I_i, W_i) = v_i(p) + bW_i - cI_i.$$

The game has four stages:

In the first stage or entry stage, each citizen simultaneously decides whether or not to enter the race and become a candidate. If no candidate enters the race, then the game ends and a default policy p_0 is implemented. If only one candidate enters the race, she is automatically elected, she implements her ideal policy, and the game ends.

If at least two candidates compete for office, in stage two or support stage, each citizen decides to support one of the candidates. Citizens who are indifferent toward all the candidates choose randomly which candidate to support.

In stage three, each citizen i who supports candidate j casts a valid vote for j with probability $(1-\mu)$. With probability $\mu \in [0,1)$, citizen i is unable to cast a valid vote for any candidate, so $i's$ support is lost. This probability μ is the same for each citizen and it is uncorrelated among citizens. The intuition is that citizen i supports j and intends to vote for j, but with probability μ, some random factor prevents citizen i from casting a valid vote; perhaps i cannot make it to the polls, or i votes but somehow the ballot is cast incorrectly and is later declared invalid and does not add to the total vote count.

Once all valid votes are counted, a winner is chosen by plurality rule: the candidate who has most valid votes (not necessarily the one with most intended support) wins, provided that she obtains at least two valid votes. If no candidate obtains more than one valid vote, the game returns to the beginning of stage three and citizens are called to vote again for the candidates they had decided to support in stage two.[3] In case of a tie for victory between two or more candidates with at least two votes, the winner is randomly determined among all candidates with most votes.

Finally, in stage four, the winner implements her ideal policy.

Citizens know the ideal policy of each candidate and they correctly anticipate that the winner will implement her ideal policy; candidates cannot commit at the entry stage to implement any other policy if they win the election. Empirical evidence such as Lee, Moretti and Butler [37] indicates that, indeed, voters merely elect the policy they prefer among those offered by the competing candidates,

[3] This mild assumption prevents a candidate winning with her own vote only and no support in the rest of the electorate, and it is technically convenient. However, if every citizen ran as a candidate and supported herself, no candidate would ever get two votes. We assume that in this unlikely scenario the game ends with a negative utility payoff that makes every agent worse off than in any other outcome.

without being able to influence how the winning candidate chooses policy once in office.

The uncertainty about the vote count in stage three captures the idea that candidates cannot anticipate the outcome of the election, because they do not have enough information about the voting population. The candidates cannot anticipate whose vote will count and whose will not, so even though they anticipate the support intentions of the whole electorate, they are still uncertain about the electoral outcome.

There are other ways to incorporate uncertainty to the model. Roemer [51] lets each candidate win with probability equal to the candidate's share of votes. This assumption may seem crude but keeps his model simple and tractable. A more elegant approach would be to model turnout, assuming that the probability that a citizen votes increases with the expected benefit for the citizen from voting. With costless voting, every rational citizen who is not indifferent about every candidate should vote, but according to Quantal Response theories (see McKelvey and Palfrey [41]) citizens with little to gain or lose are more likely to make the mistake of abstaining than those with a higher stake in the outcome of the election. I do not model turnout decisions: The parameter μ is merely a simple way to capture the uncertainty faced by the candidates, who in reality, as in the model, do not know the exact distribution of preferences of the voting population.

The strategy of each citizen i has two components: an entry strategy, determining whether to run as a candidate or not, and a support strategy, which determines for any possible set of candidates which one will citizen i support.

Let $I_i \in \{0,1\}$ denote citizen i's pure entry strategy, where $I_i = 1$ denotes entry, and let $I = (I_1, ..., I_N)$ be the pure entry strategy profile of all the citizens. The set of candidates resulting from the entry strategy I is $C \subseteq \mathcal{N}$. Let $\gamma_i \in [0,1]$ be the mixed entry strategy by citizen i, indicating the probability that i enters the race, and let γ be the entry strategy profile of all citizens.

Let $s_i : 2^N \longrightarrow \mathcal{N}$ denote the pure support strategy used by citizen i, which determines for each possible set of candidates, which one citizen i will support. For a given set of candidates C, $s_i(C)$ denotes the candidate supported by citizen i. Let $s = (s_1, ..., s_N)$ be the pure support strategy profile of every citizen in the society. Let σ_i be the mixed support strategy profile of citizen i, and let σ be the mixed support strategy profile of every citizen. Let the subscript $_{-i}$ denote "every citizen in $\mathcal{N}$ except for i."

The equilibrium concept I use is Undominated Subgame Perfect Nash Equilibrium, ruling out weakly dominated strategies. An equilibrium is defined by an entry strategy profile γ^* and a support strategy profile σ^* such that:

(i) Given any set of candidates $C \subseteq \mathcal{N}$ of size at least two, $\sigma^*(C)$ is an Undominated Nash equilibrium of the support stage subgame.

(ii) Given σ^*, the entry strategy profile γ^* is a Nash equilibrium at the entry stage of the game.

When citizens use pure entry strategies, I let C^* denote the set of candidates who run in the equilibrium $\{\gamma^*, \sigma^*\}$.

I assume that each agent i has a unique favorite policy $p_i = \arg\max_{p \in [0,1]} v_i(p)$ and I label and order individuals according to their favorite policy, so that for all $i, j \in \mathcal{N}$, $i < j$ implies $p_i \leq p_j$. Given this ordering of citizens, I assume that the preference profile satisfies the Strict Single-Crossing property:

Definition 1 *A preference profile satisfies the Strict Single-Crossing property if for all $x, y \in [0,1]$ and all $i, j \in \mathcal{N}$ such that $y > x$ and $j > i$, $v_i(y) \geq v_i(x) \Longrightarrow v_j(y) > v_j(x)$.*

This property implies that if a left-leaning citizen d prefers the right-most of two policies, then every citizen who is more right-leaning than d also prefers the right-most policy. Given two policies, one more liberal, one more conservative, there cannot be any overlapping so that "conservative" citizens support the liberal policy and some more "liberal" citizens support the conservative policy: With strictly single-crossing preferences, given any two policy positions $p' < p''$, there exists a cut-off point $\widehat{p}(p', p'')$ such that every citizen with an ideal policy below $\widehat{p}(p', p'')$ prefers p' to p'' and every citizen with an ideal policy above $\widehat{p}(p', p'')$ prefers p''.

Borrowing the terminology of the Poole and Rosenthal [46] Nominate Scores model, single-crossing preferences imply "perfect voting": The voting behavior of a citizen-voter faced with a binary choice should be perfectly explained by the position of the citizen in the unidimensional low-to-high scale. Poole [45] notes that in practice, more than 90% of the votes in US Congress can be explained by the position of a legislator in a one-dimensional policy space, although this evidence is merely suggestive when we are studying large elections.

I introduce the following notation:

Let m denote the median voter if N is odd, and let m_l and m_h denote the two medians if N is even. Labelling citizens by the relative position of their ideal policy from lowest to highest, m is the citizen in position $\frac{N+1}{2}$ if such fraction is an integer; otherwise m_l and m_h are respectively the citizens in positions $\frac{N}{2}$ and $\frac{N+2}{2}$.

Let $S_i(C, \sigma)$ denote the support for candidate i, that is, the number of citizens whose support strategy is to support candidate i, given that the set of candidates is C and the joint support strategy profile is σ. If all citizens use a pure support strategy, $S_i(C, \sigma) = \#\{j : s_j(C) = i\}$. If citizens use mixed support strategies, S_i is a random variable that can take different values depending on the support actions taken by the agents who mix. Let then $E[S_i]$ denote the expected value of S_i.

Let V_i denote the number of valid votes for i. The difference $S_i - V_i$ corresponds to the number of citizens who support i but are unable to cast a valid vote for i and can be interpreted as the number of "lost votes" for i.

In any two-candidate race with $C = \{i, j\}$, let

$$L_{ij} = (S_i - S_j) - (V_i - V_j) = (S_i - V_i) - (S_j - V_j)$$

denote the "shift" from candidate i to candidate j in the difference of voting totals for the two candidates, compared to the original difference in support for the two candidates prior to the distortion introduced by the loss of votes. In short, L_{ij} is equal to the number of votes lost by i minus the number of votes lost by j, which we can interpret as a "net loss" of votes for i.

L_{ij} is a discrete random variable, whose distribution depends on the support for each candidate and the uncertainty parameter μ: Let $f_{ij}(l)$ be its probability mass function and let $F_{ij}(l)$ be its distribution function, so $f_{ij}(l)$ is the probability that $L_{ij} = l$ and $F_{ij}(l) = \sum_{k=-N}^{l} f_{ij}(k)$.

In any race with two candidates $C = \{i, j\}$ with ideal policies $p_i < p_j$, let $\widehat{p}_{ij}$ denote the cut-off point such that every citizen with an ideal policy less than $\widehat{p}_{ij}$ prefers candidate i, every citizen with an ideal policy above $\widehat{p}_{ij}$ prefers candidate j, and only citizens with an ideal policy equal to $\widehat{p}_{ij}$ are indifferent between i and j.

2.3 Existence of Equilibria

Let us first present the benchmark case in which there no uncertainty about the vote count ($\mu = 0$) and the candidates can anticipate the outcome of the election. This benchmark corresponds to the Besley and Coate [9] model with the additional assumptions of strict single-crossing preferences in a unidimensional policy space, and $b > 2c$.

A single-candidate equilibrium exists if and only if N is odd, whereas two candidate equilibria generically do not exist if N is odd and they exist if N is even. Equilibria with multiple candidates are also possible.

These differences depending on the exact number of citizens are plausible in an election with a small electorate, such as a vote in a committee. However, in any election with a very large electorate, there will be some uncertainty about the number of votes each candidate will get and the results of the model should not depend on whether the size of the electorate is odd or even.

In the rest of the chapter, I capture the uncertainty assuming that $\mu > 0$ and I show that if the electorate is sufficiently large, whether N is odd or even is irrelevant for the existence of equilibria with one or two candidates in elections with a large reward for holding office.

If citizens use mixed entry strategies, the number of candidates who run may vary in different outcomes of the same equilibrium. I categorize equilibria according to the number of citizens who enter the race with positive probability. If citizens use only pure strategies at the entry stage, I say the equilibrium is *pure.*

Definition 2 *An n-candidate equilibrium is an equilibrium in which the number of citizens who run with positive probability is n.*

Definition 3 *An equilibrium is "pure" if every citizen uses a pure entry strategy.*

2.3.1 Single-Candidate Equilibrium

In this subsection, I characterize existence of single-candidate equilibria, and I show that they do not exist in a society with a sufficiently large electorate. In a unidimensional space, a unique median is a Condorcet winner and will have more support (and more expected vote share) than the other competitor in any two-candidate race.

Nevertheless, the uncertainty about the vote totals gives any other candidate challenging the median some positive probability of winning the election, and in equilibrium the median can only run alone if the probability of victory for any candidate running against her is too low.

Lemma 1 *There exists a one-candidate equilibrium if and only if N is odd, the median is unique, and for any citizen $j \in \mathcal{N} \backslash m$,*

$$(b + v_j(p_j) - v_j(p_m)) \Pr[W_j = 1 | C = \{m, j\}] \leq c.$$

This equilibrium is unique and pure, and m is the single candidate.

This and all other proofs are in an appendix at the end of the chapter. The intuition is that the median(s) would enter and run against any other citizen who was running alone, so only a unique median can run unopposed, and even the median can run unopposed only if any challenger would have a small enough probability of victory.

A candidate trailing by a small number of supporters almost certainly loses if the electorate is also small, but as the electorate gets larger, the number of lost votes will increase and a candidate trailing by the same small number of supporters will have a better chance of victory. For instance, in an electorate with 5 citizens, a $3-2$ split of support will give the weaker candidate a very small chance of victory. However, in an electorate with millions of citizens, a split of support in which the stronger candidate has only one more supporter is a virtual tie, and both candidates have an almost equal probability of victory.

Given a fixed parameter of uncertainty μ, consider a sequence of societies of increasing size. Consider also a corresponding sequence composed of a pair of candidates in each society such that the difference in support for the two candidates in each pair is constant along the sequence. For example, in society $\mathcal{N}_N$ of size N construct the pair with the median m_N and the next citizen, $(m+1)_N$; regardless of the distribution of preferences in each society, for every element of this sequence of pairs the median has one more supporter than the candidate in position $m+1$. The probability of victory converges to a half for both candidates as the size of the society increases. I use this result for the first theorem on existence of a single-candidate equilibrium in large societies.

Theorem 2 *Given $\mu > 0$, there exists some n such that if $N > n$, there is no single-candidate equilibrium.*

Suppose the benefit of holding office is three times the cost of running. Then any candidate with a one-third chance of victory will be willing to run. The probability of victory for a weaker candidate with one less supporter than the median is more than a third if $N \geq 103$ for $\mu = 0.05$; or if $N \geq 1087$ for $\mu = 0.005$. Since void ballots exceed 0.5% in most elections, these numerical examples show that the theorem applies for relatively small electorates. A higher degree of uncertainty about counting an individual vote has the same effect as increasing the size of the population: In either case it becomes harder for the agents to anticipate the exact outcome of the election and a challenger will have greater incentives to run against the median.

In an election with a small electorate the median can run unopposed because any other candidate would have only a very slim chance of beating the median. The median must be unique to run unopposed, so it is crucial that the electorate is odd. However, as the electorate gets larger, the uncertainty about the vote count gives other candidates challenging the median a better chance of winning, and for a sufficiently large society the probability of victory for citizens with an ideal point close to the median is high enough so that they will run and not let the median win unopposed, not even a unique, Condorcet winner median. Note that this result does not at all depend on the distribution of preferences in the population: As long as the society is large enough, some other citizen will challenge the median.

As a result, in any equilibrium at least two citizens will enter with positive probability. If citizens use mixed entry strategies, it could be that in a particular outcome only one candidate stands for election, but the positive probability of entry by other citizens must be part of the equilibrium.

2.3.2 Two-Candidate Equilibria

Two-candidate equilibria are a common feature of plurality elections, yet they are generically non-existent in citizen candidate models without uncertainty if the number of citizens is finite and odd. I show that introducing uncertainty in the model guarantees the existence of two-candidate equilibria if the size of the society is sufficiently large.

To obtain this result, I add a mild restriction on the preference profile; I require that each citizen has a distinct ideal policy. The formal condition is that for any $i, j \in \mathcal{N}$, $p_i \neq p_j$.

Intuitively, this assumption is necessary to avoid entry by multiple candidates with the same policy. As an illustrative extreme case, suppose all citizens but one share a common ideal policy, and suppose that the cost of running is negligible compared to the benefit of holding office. Then all the citizens with the same ideal policy run against each other. No equilibria with less than all of them running is possible. However, if the rest of the electorate is able to discriminate among any pair of candidates, then in a large enough electorate, equilibria with two candidates exist.

Theorem 3 *Given $\mu > 0$, there exists some n such that if every citizen has a distinct ideal policy* *and $N > n$, a pure two-candidate equilibrium exists.*

The proof is constructive. If there is a unique median, in the absence of uncertainty (or with little uncertainty in small electorates), two-candidate equilibria did not exist unless the median citizen is indifferent between the two candidates, an event that generically does not occur.

However, as the electorate grows, a positive uncertainty raises the probability that a candidate trailing by one supporter wins the election. If there are large benefits of holding office, the same intuition that made single-candidate equilibria impossible guarantees the existence of two-candidate equilibria for large electorates. As the electorate grows, the probability that a weaker candidate trailing by a given number of supporters wins the election converges to one-half, and $b > 2c$ guarantees that such a weaker candidate will want to run, and that equilibria with two candidates exist in large societies, whether the number of citizens is even or odd.

Two-candidate equilibria with mixed entry strategies may also exist,[4] but only the weaker candidate will mix: The stronger candidate has a probability of victory which is over one half, so she prefers to enter. Furthermore, if there is a two-candidate equilibrium in which candidate i enters for sure and candidate j mixes between running and not running in her entry decision, there is also a pure equilibrium in which both i and j enter with probability one.

Proposition 4 *If there exists a two-candidate equilibrium in which i and j run with positive probability, then there exists a pure two-candidate equilibrium in which i and j run.*

Only a weaker candidate with not so good odds of victory may be indifferent about running or not. The weaker candidate may only mix if she is indifferent about running. If she is indifferent, it is also a best response for her to run with probability one. The stronger candidate wants to run regardless of the probability of entry by the weaker one. If both candidates have the same support, both enter with probability one.

[4] Previous literature has paid only scant attention to equilibria with mixed entry. In a novel experimental paper, Cadigan [13] provides an example.

Since considering mixed entry does not expand the set of pairs of citizens who may run in a two-candidate race, I focus on equilibria which are pure at the entry stage. In the constructive proof of theorem 3, the median(s) or a citizen very close to the median are the two candidates who enter for sure in equilibrium. However, there exist other equilibria involving citizens far from the median: Two very extreme candidates i and j can run in equilibrium, insofar as the cutting point $\widehat{p}_{ij}$ between those who support i and those who support j is very close to the median, (or to the medians if N is even).

Let i and j be the any two citizens, such that $p_i < p_j$. Let $q(x)$ be the minimum number of supporters that i must have in order for i to win with probability no less than x, given that every other citizen will support j. That is, $q(x)$ is the minimum k such that:

$$\Pr[W_i = 1 | S_i = k, S_j = N - k] \geq x.$$

In the following proposition I use the function q to specify the number of supporters each candidate must have in order to sustain a two-candidate equilibrium.

Proposition 5 *Suppose each citizen has a distinct ideal policy. For any $i, j \in \mathcal{N}$ such that $p_i < p_j$, let*

$$D = q\left(\frac{c}{b + v_i(p_i) - v_i(p_j)}\right) \text{ and } R = q\left(1 - \frac{c}{b + v_j(p_j) - v_j(p_i)}\right), \ D, R \in \mathcal{N}.$$

If $\widehat{p}_{ij} \in (p_D, p_R)$, there exists a pure two-candidate equilibrium in which i and j run.

There is a two-candidate equilibrium if both candidates get a similar number of supporters and thus they both have a sufficiently high probability of victory. In order for the electorate to split in roughly equal halves, the cutting point $\widehat{p}_{ij}$ between those who support i and those who support j must be close to the median. Proposition 5 shows it suffices that the cutting point lies in between the ideal policies of citizens D and R, where the function $q(x)$ specifies the identity of D and R, which depends on the identity of i and j.

A very slight weakening of the sufficient condition in proposition 5 is already a necessary condition: If i and j run in a pure two-candidate equilibrium, then $\widehat{p}_{ij} \in [p_D, p_{R+1})$. Only equilibria in which one of the candidates is indifferent about running or not need not satisfy the sufficient condition.

Since $b > 2c$, each candidate is willing to run for a probability of victory less than a half and it follows that D is weakly to the left of the median m if N odd, or of the low median m_l if N is even. Similarly, H is weakly to the right of m or m_h. Suppose N is even. Then if m_l and m_h have different ideal policies, $\widehat{p}_{m_l,m_h} \in (p_{m_l}, p_{m_h}) \subseteq (p_D, p_R)$ and there exist a two-candidate equilibrium with pure entry strategies in which m_l and m_h run against each other.

Corollary 6 *Suppose each citizen has a distinct ideal policy. If the number of citizens is even, a pure two-candidate equilibrium exists.*

In an equilibrium without uncertainty, the two candidates must have equal support. With uncertainty, they must have similar, not necessarily equal support. In larger societies, the margin by which the weaker candidate trails in support may be bigger in absolute terms, but the fraction of the population that supports each candidate must converge to a half in a two-candidate equilibrium as the electorate gets larger. If the weaker candidate lags heavily in support, she would abandon the race.

The cutting point $\widehat{p}_{ij}$ must be close to the median in order for the candidates to have similar support. But the candidates themselves need not be close to the median. There is no convergence result in terms of the policy that will ultimately be implemented, but only in terms of the "undecided voter," the citizen who is indifferent between the two candidates. This citizen (if it exists) ought to be close to the median, splitting the electorate into two halves of roughly the same size, so that both candidacies are competitive.

The two candidates can be two moderates, or two extremists (one from each extreme), or anything in between so long as they split society into two groups of similar size.

2.4 Sincere-Strategic Equilibria

To address the multiplicity of two-candidate equilibria found in the previous section, I select equilibria according to their robustness to the assumption on the citizens' support behavior. I introduce a new concept of equilibrium, which allows for both strategic and sincere support behavior.

Definition 4 *A sincere-strategic equilibrium is a pair* $\{\gamma^*, \sigma^*\}$ *such that:*

(i) The support strategy profile σ^* *is an Undominated Nash Equilibrium of the support stage for any set of candidates* $C \subseteq \mathcal{N}$.

(ii) The entry strategy profile γ^* *is a Nash Equilibrium at the entry stage given any subset* $E \subseteq \mathcal{N}$ *such that at the support stage every citizen* $i \in E$ *follows the strategy* σ_i^* *and every citizen* $j \notin E$ *follows a sincere support strategy.*

In a sincere-strategic equilibrium, γ^* must be the equilibrium strategy of the entry stage regardless of whether every citizen acts strategically or sincerely in the support stage. Furthermore, γ^* must also be the equilibrium strategy of the entry stage if any subset of agents deviate from the strategic equilibrium and choose whom to support sincerely.

Published estimates of the incidence of strategic voting behavior in the US and the UK (two countries that elect a single representative per district) range between 5% and 17%, as reported by Alvarez and Nagler [3]. A sincere-strategic equilibrium is a Strategic equilibrium that is robust at the entry stage to deviations by any coalition of agents toward sincerity at the support stage. While it is difficult for citizens to coordinate strategically, it is easier to hypothesize how they might be convinced as a group to vote sincerely: An appeal by a candidate simply to the "honesty" of the citizens, to "vote according to your heart" might convince a large number of voters to deviate from the equilibrium strategic behavior more readily than a complicated appeal to coordinate on a sophisticated deviation. The refined equilibrium concept requires the equilibrium to be robust to successful appeals to sincerity.

It is trivial to note that every single-candidate equilibrium is a sincere-strategic equilibrium. In a single-candidate equilibrium, citizens automatically elect the only candidate and if a second candidate entered the race, with two candidates strategic support is equivalent to sincere support. Therefore, whether citizens choose whom to support sincerely or strategically does not alter the entry decision of any citizen when there is a single candidate who runs with positive probability.

However, if there are two candidates running, citizens who act sincerely may support an entrant who would have been shunned if all citizens acted strategically.

In particular, note that a two-candidate equilibrium with two extreme candidates cannot be a sincere-strategic equilibrium, because a moderate entrant could gather a significant amount of support, enough to have a high enough probability of victory. Thus some convergence is necessary for a sincere-strategic equilibrium.

Nevertheless, full convergence is not possible. Two adjacent candidates would also fail to stand in a two-candidate sincere-strategic equilibrium, because another adjacent candidate, just a little bit more extreme than either of the original two can enter and outflanking one of the candidates, get the support of almost a half of the electorate. For example, if m_l and m_h were running, then $m_h + 1$ can run, and under sincere support behavior, get the support of $\frac{N}{2} - 1$ citizens, leaving m_l with $\frac{N}{2}$ supporters and m_h with just his own support.

Remark 7 *Sincere-strategic equilibria with two candidates consist of a left-leaning candidate and a right-leaning candidate, satisfying three conditions:*

(i) They are both separate from the median,

(ii) they are not too far from each other,

(iii) and they split society into two groups with roughly the same number of supporters.

The candidates have to be separate from the median to make an extreme third candidacy inviable: If a third candidate entered to the left of the left-leaning candidate, the right-leaning candidate would win if enough leftist citizens were sincere and the left vote is split, or else the entrant would have no support and would not affect the election (if citizens are strategic).

They cannot be too far from each other (too extreme) or else a moderate third candidate would enter in between them and the entrant would sweep the election if citizens chose support sincerely.

The two candidates have to split society into two roughly equal groups of support to be both viable candidates with a good chance of winning the election.

The following example illustrates these three features of sincere-strategic equilibria.

Example 1 *Let there be 100 citizens, with ideal policies* $p_i = \frac{i}{100}$ *for all* $i \in \{1, 2, ..., 100\}$. *Let the policy preferences be:*

$$v_i(p) = \left\{ \begin{array}{c} -0.1(p_i - p) \text{ if } p \leq p_i. \\ \\ -0.101(p - p_i) \text{ if } p > p_i \end{array} \right\},$$

so every citizen i *prefers the candidate who is closest to* i *and in case of equal distance, the candidate with the lower ideal policy. Let uncertainty* $\mu = 0.025$ *and let* $b = 10c = 10$.

Then, in any two candidate sincere-strategic equilibrium with $C = \{i, j\}$*:*

Either $(i, j) \in \{(47, 53), (48, 53)\}$ *or* $i \in \{17, 18, ..., 47\}$, $j \in \{54, 55, ..., 83\}$, $\frac{(i+j)}{2} \in [49, 51.5]$, *and* $|i - j| \leq 64$.

I provide the calculations in the appendix to the chapter.

Next I show a set of sufficient conditions that guarantee the existence of a pure sincere-strategic equilibrium with two candidates in a large electorate. These conditions include Euclidean preferences with two mild but technical restrictions on the distribution of ideal policies, which has to be somewhat evenly spread and not too clumped to one side of the median, and a simpler but more restrictive assumption on the cost of running as a candidate, which I require to be high enough to deter entry by candidates with no chance of winning.

Let η be the maximum amount of support by which a candidate i could be trailing in a two-candidate race and still be motivated to run solely for the expected benefit of holding office. Formally, let

$$\eta = \left| \min\{l : \frac{b}{2} f_{ij}(l) + bF_{ij}(l-1) > c\} \right|$$

and let E_η be the smallest convex interval containing the ideal policies of at least η citizens.

Definition 5 *The distribution of ideal policies is "nonclumped" if any convex interval $E \subseteq [0,1]$ of weakly greater length than E_η contains the ideal policy of at least one agent.*

An informal interpretation of "nonclumped" preferences is that ideal policies are somewhat spread all over the policy space, and there are no big empty gaps coexisting with thin intervals where a lot of ideal policies lie in a cluster.

I also require the distribution of ideal policies to be balanced near the median, and not too lopsided with a lot of ideal policies within a small distance to one side of the median, and very few within the same distance to the other side of the median.

Let $\lceil x \rceil$ denote the smallest integer equal or larger than x. I use this notation in the following definition.

Definition 6 *For any $\varphi \in [0,1]$, the distribution of ideal preferences is "$\varphi -$ balanced" if p_m is weakly closer to $p_{m-\lceil \frac{\varphi}{8}N \rceil}$ and $p_{m+\lceil \frac{\varphi}{8}N \rceil}$ than to either $p_{\lceil \frac{3}{8}\varphi N \rceil}$ or $p_{N+1-\lceil \frac{3}{8}\varphi N \rceil}$.*[5]

Preferences are $\varphi -$ *balanced* if, moving away from the median simultaneously in both directions, before we find the ideal policies of $\frac{4-3\varphi}{8}N$ agents in the direction more dense with ideal policies, we must have also found the ideal policies of at least $\frac{\varphi}{8}N$ agents in the direction more sparse in ideal policies. The definition is increasingly milder as φ becomes smaller. Informally, for very small φ, preferences are $\varphi -$ *balanced* if moving away from the median we find at least a few ideal policies in the direction less dense with ideal policies before we find almost all the ideal policies that lie to the side of the median more dense in ideal policies. For example, let $N = 1001$. Then preferences are

[5] *For an even N, we adapt the definition as follows:*
The distribution of ideal preferences is "$\varphi -$ balanced" if $\frac{p_{m_l}+p_{m_h}}{2}$ is weakly closer to $p_{m_h-\lceil \frac{\varphi}{8}N \rceil}$ than to $p_{N+1-\lceil \frac{3}{8}\varphi N \rceil}$ and $\frac{p_{m_l}+p_{m_h}}{2}$ is also weakly closer to $p_{m_l+\lceil \frac{\varphi}{8}N \rceil}$ than to $p_{\lceil \frac{3}{8}\varphi N \rceil}$.

$1\% - balanced$ if p_{499} and p_{503} are both closer to the median ideal policy p_{501} than either p_4 or p_{998}. At the extremes, if p_m is equidistant to $p_{\frac{3}{8}N}$ and $p_{\frac{5}{8}N}$, then preferences are $1 - balanced$; whereas, if p_m is closer to p_N than to p_{m-1}, then preferences are $0 - balanced$.

Euclidean preferences are defined by the Euclidean distance between the ideal policy of the agent and the implemented policy: $v_i(p) = |p - p_i|$.

Finally, let $\overline{c} = \max\limits_{i,j,h \in \mathcal{N}} |v_i(p_j) - v_i(p_h)|$. The following proposition details the existence result.

Proposition 8 *Given any* $\mu, \varphi > 0$, *there exists some* n *such that if preferences are Euclidean, nonclumped and* $\varphi - balanced$, *each citizen has a distinct ideal policy,* $c > \overline{c}$ *and* $N > n$, *then a pure sincere-strategic Equilibrium with two candidates exists.*

The proof in the appendix is constructive: It finds two citizens who are willing to run against each other, it shows that no third agent has a chance of victory if citizens choose support sincerely, and it uses the assumption of a high cost of running to rule out entry by candidates with no chance of victory.

2.4.1 Sincere-Strategic Equilibria with 3 or More Candidates

Equilibria with multiple candidates, and sincere-strategic equilibria with multiple candidates are possible in the model. In this subsection I show that under some restrictions, the sincere-strategic refinement eliminates all equilibria in which weak candidates run with little support.

Models without uncertainty can distinguish between "winning candidates" (those who win with positive probability) and "spoilers" (those who run just to affect the outcome indirectly, but with no probability of actually winning). With uncertainty, every candidate with at least one supporter has some positive probability of victory, albeit possibly a very small one. Nevertheless, we can still describe some candidates as spoilers: Those who trail in support and run mostly to influence who wins when they lose.

Formally, I define competitive candidates as those who have enough of an incentive to run based just on their probability of winning, without taking into account the effect that their candidacy would have on the electoral outcome if they lose. Competitiveness is defined relative to a specific joint strategy of all citizens at the support stage; thus a candidate can be competitive for some profile of support and not competitive given a different support profile.

Definition 7 *Given a joint support strategy profile σ, a candidate $i \in C$ is competitive if*

$$\{b + v_i(p_i) - \sum_{k \in C \backslash i} v_i(p_k) \Pr[W_k = 1 | C \backslash i]\} \Pr[W_i = 1 | C] > c,$$

and a spoiler otherwise.

Competitive candidates run to win. Spoiler candidates run motivated by the advantages that running for a defeat entails for them. Equilibria with three competitive candidates, which do not exist under mild assumptions in Besley and Coate's [9] model, exist in this model with uncertainty (an example is available from the author). Equilibria with four candidates may have one or two competitive candidates, as in Besley and Coate [9], or three or four, as in Osborne and Slivinski [44]. However, with the refinement of the sincere-strategic equilibrium, in all equilibria in which citizens use pure strategies at the entry stage, spoilers must all be in between two competitive candidates. Therefore, pure equilibria with a single competitive candidate and several spoilers will not be sincere-strategic.

Lemma 9 *In any pure sincere-strategic equilibrium with multiple candidates , if $i \in C^*$ is a spoiler, then there exist $j, h \in C^*$ such that $p_i \in (p_j, p_h)$.*

Corollary 10 *In any pure sincere-strategic equilibrium with two or more candidates, there exist at least two competitive candidates.*

Entry by spoiler candidates who deprive stronger candidates from crucial support can effectively affect the outcome of an election. However, the entry decision by a candidate who draws no support from the electorate is intuitively irrelevant to the outcome of an election. If a candidate with no support drops out of the race, the most intuitive reaction by all other citizens is to keep their support decisions unaltered.

I say that a support strategy profile is *consistent* if whenever a candidate who receives zero support drops out of the race, the support for all the other candidates remains unaltered.

Definition 8 *A support strategy profile* σ *is consistent if for any* C *and* $i \in C$, $S_i(C, \sigma) = 0 \Longrightarrow \sigma_j(C) = \sigma_j(C \backslash i)$ *for all* $j \in \mathcal{N}$.

Next I show that if preferences are Euclidean, support strategies are *consistent,* and the electorate is large enough, then in equilibrium all candidates must have a similar expected share of support. For a fixed μ, I consider a sequence of societies of increasing size N and I show that the result holds for all societies larger than some size n. To prove this I add a technical assumption on the distribution of ideal policies to guarantee that no agent is indifferent between two distinct ideal policies.

Let $Q \subset [0,1]$ be a set composed of finitely many points such that for any $x, y \in Q$, $\frac{x+y}{2} \notin Q$.

Given a positive uncertainty μ and a set Q composed of finitely many staggered points, if preferences are Euclidean, the ideal policy of every agent lies in the set Q, and the electorate is large enough, then in any pure sincere-strategic equilibrium with *consistent* strategy profiles all candidates receive a similar expected share of support. Proposition 11 states this result formally.

Proposition 11 *Given* $\mu > 0$ *and* Q, *for any* $\epsilon > 0$ *there exists a positive integer* n *such that if preferences are Euclidean,* $p_i \in Q$ *for all* $i \in \mathcal{N}$, *and* $N > n$, *then in any pure sincere-strategic equilibrium with consistent strategy profiles*

$$\left| \frac{E[S_i]}{N} - \frac{E[S_j]}{N} \right| < \epsilon \text{ for all } i, j \in C^*.$$

The intuition for the proof is as follows: If the electorate is large enough, a candidate whose expected share of support is less than that of the strongest candidates will lose with probability approaching one. Then, by lemma 9 and corollary 10, it must be that this weak candidate is a spoiler, between two competitive candidates. But then, given two competitive candidates and a trailing weak candidate, votes for the spoiler are wasted votes, and only citizens who are indifferent among the top two candidates will support the spoiler. Indifference between two candidates with distinct ideal policies is ruled out by assumption and as a result the spoiler will get no support, then the spoiler not only cannot win, but with zero support she cannot even influence the election if the support strategy profile is *consistent*. Then, she prefers to drop out.

A complete characterization of multi-candidate equilibria is beyond the scope of this thesis.

2.5 Conclusion

I have introduced a model of representative democracy with endogenous candidates and uncertainty about the total vote count. I predict that the median is able to run a successful, unopposed campaign only if the number of citizens is small and odd. If the electorate is large, the exact number of citizens is irrelevant; no citizen can run unopposed and a two-candidate equilibrium exists.

I have characterized equilibria with one and two candidates in pure and mixed strategies. I have found that only candidates who run in some pure equilibrium may also run in a mixed equilibrium.

I have introduced a refined equilibrium concept, sincere-strategic equilibrium, which requires the equilibrium of the entry stage of the game to be robust to coalitional deviations towards sincerity by any subset of agents at the support stage, when citizen choose which candidate to support in the election. I show that in a sincere-strategic equilibrium with two candidates, the candidates cannot be too extreme, but one will lean to the left and the other one to the right; thus, policies do not converge to the median. I showed a set of sufficient conditions for this equilibrium to exist.

If the strategies used by the agents satisfy a consistency requirement, the refined equilibrium concept rules out all multi-candidate equilibria in which some candidates run with little support. If equilibria with three or more candidates exist, then all candidates must have a similar share of support.

I now compare my results with those of the two most prominent models of representative democracy with endogenous candidates: Osborne and Slivinski's [44] and Besley and Coate's [9].

Besley and Coate [9] consider a finite number of strategic citizens. Candidates perfectly anticipate the outcome of the election, which depends crucially on the exact (odd or even) number of citizens: Assuming that the benefit of holding office is large, single-candidate equilibria exist if the median is unique (or if there exist a Condorcet winner if the policy space is not unidimensional) and two candidate equilibria exist if the number of citizens is even. This model best fits an election with a very small electorate and with such an electorate, my model yields similar results. The key difference between my model and Besley and Coate's [9] is that with uncertainty about the vote count, a candidate trailing by a small margin in terms of expected votes has a high enough probability of victory and therefore this weaker candidate will choose to run.

My result for large electorate reinforces the findings of Osborne and Slivinski [44], who reach a similar prediction under a number of simplifying assumptions (sincere voting, a continuum of voters, Euclidean preferences). Their model is bound to fit large electorates better than small ones, for it relies on the existence of an infinite number of voters.

My model with a finite number of voters thus builds a bridge between the two main models of representative democracy with endogenous candidates, agreeing with each of the two models where it is most appropriate -with Besley and Coate [9] for small electorates, with Osborne and Slivinski [44] for large ones. I blend their different predictions into a single, unifying framework that captures the insights of large electorates where complete information is implausible as well as those of smaller electorates where the outcome is easier for all agents to foresee.

2.6 Appendix

2.6.1 Proof of Lemma 1

Proof. First, I prove the necessary condition. Suppose N is even and citizen $h \geq m_h$ runs unopposed. In equilibrium, support strategies for any race with two candidates are sincere. Then, if m_l enters the race and $p_{m_l} \neq p_h$, $S_{m_l} \geq S_h$ and the probability that m_l wins is at least one-half, while if $p_{m_l} = p_h$, every citizen mixes support with equal probability and m_l wins with probability one-half. Thus, $b > 2c$ implies that m_l in any case would prefer to run, and h running alone cannot be an equilibrium. Suppose $d \leq m_l$ runs unopposed. Then, by an analogous logic, m_h would prefer to run, and it cannot be an equilibrium that d runs unopposed. Therefore, there cannot be a one-candidate equilibrium if N is even.

Suppose N is odd and the median is not unique and suppose citizen $h \in \mathcal{N}$ runs unopposed. Then, if any of the medians runs against h, the median wins with probability no less than a half, so the median prefers to run, and it cannot be an equilibrium in which h runs unopposed.

Suppose the median is unique and runs unopposed. If $\exists\ j \in \mathcal{N} \backslash m$ such that

$$(b + v_j(p_j) - v_j(p_m)) \Pr[W_j = 1 | C = \{m, j\}] > c,$$

then j would run against m.

Second, I prove the sufficient condition. If a unique median m runs and

$$(b + v_j(p_j) - v_j(p_m)) \Pr[W_j = 1 | C = \{m, j\}] \leq c\ \forall j \in \mathcal{N} \backslash m,$$

then no citizen j has an incentive to run against m, and since $b > c$, m prefers to run and the entry condition of the equilibrium is satisfied. To complete the equilibrium, I only need to construct a support equilibrium σ^* for each possible support subgame reached after any off-equilibrium outcome of the entry game. Since the number of players is finite and their strategy set is also finite (there are only a finite number of candidates to choose from), such an equilibrium exists for each subgame, possibly in mixed support strategies.

To show uniqueness, it suffices to note that m would prefer to run against any other citizen $h \in \mathcal{N}$ who was running alone. The equilibrium is pure because since $b > c$, m strictly prefers to run than to drop out, given that no other citizen enters the race. ■

2.6.2 Proof of Theorem 2

Proof. The only single-candidate equilibrium is that in which a unique median is running (lemma 1). Consider a sequence of societies $\{\mathcal{N}_N\}_{N=3}^{\infty}$ of size N odd, all of them with $\mu_N = \mu$ and satisfying all the assumptions in the model. Now suppose in each society $\mathcal{N}_N$ the median and the citizen immediately to her right run, so $C_N = \{m_N, (m+1)_N\}$, where the subindex denotes the society to which a citizen or a set of candidates belongs. Then in the equilibrium of the support subgame every citizen chooses support sincerely and $S_{m_N} - S_{(m+1)_N} = 1$. If m_N loses at least three votes more than $(m+1)_N$, $V_{(m+1)_N} - V_{m_N} \geq 2$ and $(m+1)_N$ wins the election. The probability that m_N loses at least three votes more than $(m+1)_N$ is:

$$1 - F_{m,m+1}(2).$$

Let $bi[n, p; k]$ be the probability that a binomial distribution with parameters (n, p) takes a value of k. Then

$$f_{m,m+1}(l) = \sum_{k=0}^{N} bi\left[\frac{N+1}{2}, \mu; k+l\right] bi\left[\frac{N-1}{2}, \mu; k\right].$$

As $N \to \infty$, $f_{m,m+1}(l)$ converges to zero for any given integer l, in particular for $f_{m,m+1}(2)$, $f_{m,m+1}(1)$ and $f_{m,m+1}(0)$, so $F_{m,m+1}(2) - F_{m,m+1}(-1)$ converges to zero. Since $S_{m_N} > S_{(m+1)_N}$, the probability that the m_N loses less votes than $(m+1)_N$ is less than a half, so $F_{m,m+1}(-1) < \frac{1}{2}$ for all N. But $F_{m,m+1}(2) > \frac{1}{2}$ for all N, thus it must be $F_{m,m+1}(2)$ converges to $\frac{1}{2}$. Then, the probability that m_N loses at least three more votes than $(m+1)_N$ converges to $\frac{1}{2}$ and the probability that $(m+1)_N$ wins also converges to a half. Given $b > 2c$, this implies that if N is large enough, $(m+1)_N$ wants to run against m_N, and m_N cannot be the only citizen running with positive probability in equilibrium. ■

2.6.3 Proof of Theorem 3

Proof. If N is odd, I construct a pure equilibrium in which γ^* is such that $I_m^* = I_{m+1}^* = 1$, $I_j^* = 0$ for all $j \in \mathcal{N} \setminus \{m, m+1\}$, and σ^* is such that for any $C = \{m, m+1, h\}$ with $h \in \{\emptyset \cup \mathcal{N} \setminus \{m, m+1\}\}$, $s_i(C) = m$ for any $i \leq m$ and $s_i(C) = m+1$ for any $i > m$.

Given σ^*, no citizen would support a third candidate and therefore no third candidate wants to enter. Consider again a sequence of societies $\{\mathcal{N}_N\}_{N=3}^{\infty}$ indexed by their size, and for each society $\mathcal{N}_N$ select the pair $\{m_N, (m+1)_N\}$. Since $S_{m_N} - S_{(m+1)_N} = 1$ for all N, it follows that as $N \to \infty$ the probability of victory converges to one half for both m_N and $(m+1)_N$. Since $b > 2c$, m_N and $(m+1)_N$ want to run against each other if the electorate is N is large enough. Therefore, γ^* is an equilibrium of the entry stage given σ^*.

Dropping the subindex, for any society $\mathcal{N}$, the support strategy σ^* is such that citizens choose support sincerely among m and $m+1$, and that no citizen would support a third entrant. If an agent deviates from this support strategy to support a third candidate h, then $S_h = 1$; since victory in an election requires two votes, the probability of victory for h is zero, same as before the deviation. Since m and $m+1$ have a different ideal policy, it is then an undominated best response of each citizen not to support the entrant, and to choose support sincerely for either m or $m+1$, as dictated by σ^*. For any other set of candidates off the equilibrium path, the subgame at the support stage is finite and an equilibrium of the subgame exists.

If N is even, then m_l and m_h would run against each other in equilibrium regardless of the size of the society, for they each have a one-half probability of victory. With a support strategy that assigns no support to an entrant, no third candidate will enter, by the same arguments as in the odd case. ■

2.6.4 Proof of Proposition 4

Proof. Suppose there exists a two-candidate equilibrium $\{\gamma^*, \sigma^*\}$ in which $\gamma_i^*, \gamma_j^* > 0$. Without loss of generality, suppose that given σ^*, the probability that i wins given that $C = \{i, j\}$ is at least one half. Then, $b > 2c$ implies that at the entry stage, running is the only best response of i. Thus, it must be that $\gamma_i^* = 1$. Suppose $p_i = p_j$. Then the probability of victory is at least one-half for both i and j and it must be that $\gamma_j^* = 1$ too and $\{\gamma^*, \sigma^*\}$ is a pure equilibrium.

Suppose instead that $p_i \neq p_j$. Let $\gamma' = \{\gamma_j', \gamma_{-j}^*\}$, with $\gamma_j' = 1$ and let σ' be such that for any C containing $\{i, j\}$ and any $h \in \mathcal{N}$,

$$s_h'(C) = \underset{k \in \{i,j\}}{\arg\max} v_h(p_k),$$

and for any other set of candidates not including i and j, σ' defines an equilibrium of the support subgame. I want to show that $\{\gamma', \sigma'\}$ is an equilibrium. First, note σ' defines an equilibrium of any support subgame in which $\{i, j\} \in C$ because $p_i \neq p_j$ and no candidate with a single supporter can win the election (see proof of theorem 3). Second, note that $\sigma'(C) = \sigma^*(C)$ for $C = \{i, j\}$. By assumption, given σ^* and γ_{-j}^* , $\gamma_j^* > 0$ was a best response for j at the entry stage. It must then be that $\gamma_j' = 1$ is also a best response given σ^* and γ_{-j}^*, and given σ' and γ_{-j}'. I assumed that given σ^*, the probability that i wins given that $C = \{i, j\}$ is at least one half. Thus, given σ' and γ_{-i}', it is still a best response for i to run with probability one. Therefore, a pure strategy of always entering by i and j, and no entry by any other candidate is an equilibrium of the entry stage given σ' and $\{\gamma', \sigma'\}$ is a pure two-candidate equilibrium in which $C^* = \{i, j\}$. ■

2.6.5 Proof of Proposition 5

Proof. Sufficient Condition: If $\widehat{p}_{ij} > p_D$, at least D citizens support i. Then i has a probability of victory no less than $\frac{c}{b+v_i(p_i)-v_i(p_j)}$ and i wants to run. If $\widehat{p}_{ij} < p_R$, candidate i has a probability of victory less than $1 - \frac{c}{b+v_j(p_j)-v_j(p_i)}$, and citizen j wants to run. If both i and j want run, and given that they have different ideal policies, there exist an undominated support strategy that would assign no support to any third entrant, and then no third citizen would want to deviate and enter.

$\blacksquare$

I also prove the necessary condition enunciated in the text:

Proof. If $\widehat{p}_{ij} < p_D$, then at most $D-1$ citizens support i, and at least $N-L+1$ support j. Then, the probability of victory for citizen i is less than $\frac{c}{b+v_i(p_i)-v_i(p_j)}$ and i prefers to let j run alone and win. If $\widehat{p}_{ij} \geq p_{R+1}$, at least R citizens support i and $R+1$ at least mixes (or supports i too), and at most $N-R-1$ support j, with citizen $R+1$ at most mixing. Then, the probability that i wins is more than $1 - \frac{c}{b+v_j(p_j)-v_j(p_i)}$, and citizen j does not want to run. $\blacksquare$

2.6.6 Calculations for Example 1

Citizens will want to run if their chance of victory is better than 10% just out of motivation to hold office, and depending on who else is running, they may want to run for a chance of victory slightly less than 10%, out of policy considerations. No citizen wishes to run for a probability of victory less than 9%.

If the two top candidates split support $49-51$, $48-50$, or $47-49$, the weaker candidate has more than 10.5% probability of victory, thus she is willing to run, but with a split of $48-51$ or worse (trailing by three or more supporters), the weaker candidate has a probability of victory below 5% and she would not want to run. With just two candidates i and j, citizens split support $49-51$ if $i+j$ adds up to 98, 99 (in which case i gets support from citizens 1 to 49), 102 and 103 (in which case citizens 50 and 51 support i as well). Citizens split support $50-50$ if $i+j$ adds up to 100 or 101.

Thus in order for i and j to be willing to run against each other, it must be $98 \leq i+j \leq 103$.

However, if j is 50, or 51, then a third candidate $h = 52$ could enter and get 49 supporters (if citizens happen to be sincere), enough to make h competitive. If $j = 52$ and $i = 49, 50$ or 51, then

$d = 48$ could enter and get at least 48 supporters under sincere support, leaving i with 1 or 2, and j with no more than 50. Then d would enter. If $j = 52$ and $i < 49$, then $h = 53$ could enter and get 48 supporters, leaving i with no more than 50 and j with 2 or 3. Then h would run.

Therefore, it cannot be that j is less than 53. And if $j = 53$, then i cannot be less than 47, or else $h = 54$ with 47 supporters would enter and face i with no more than 49 (j would keep the support of citizens 50, 51 and 52).

Similarly, if i is 49 or 50, then $d = 48$ could enter and get 48 votes, enough to be competitive against the (no more than 50) votes of j.

If $i = 48$ and $j = 53$, then outflanking is no longer viable for a third candidate: If $d = 47$ enters, then d gets 47 votes and j gets 50.

So besides adding up to 98, 99, 100, 101, 102 or 103, it must be that either $i, j \notin \{48, 49, 50, 51, 52, 53\}$ or $(i, j) \in \{(47, 53), (48, 53)\}$.

Finally, i and j cannot be too far apart, or else a moderate citizen will enter:

If $j - i$ is more than 64, then either $m_l = 50$ or $m_h = 51$ could enter and get no less than 33 supporters and leave the other two candidates with 33 and 34 supporters (or 34 and 33). In such a tight three-way race, the probability that the weakest candidate wins the election is over 10%.

2.6.7 Proof of Proposition 8

Proof. For a sufficiently large N, I construct an equilibrium in which $I_i^* = I_j^* = 1$ and $I_h^* = 0$ for all $h \in \mathcal{N} \backslash \{i, j\}$, and σ^* is such that for any set of candidates that includes i and j, all citizens support either i or j, and for any other set of candidates (off the equilibrium path) an arbitrary equilibrium of the subgame is played.

Without loss of generality suppose that p_m is closer to $p_{m+\lceil \frac{\varphi}{8}N \rceil}$ than to $p_{m-\lceil \frac{\varphi}{8}N \rceil}$. Then let $i = m - \lceil \frac{\varphi}{8}N \rceil$.

Now construct the interval $X = [p_m + p_{m-\lceil \frac{\eta}{2} \rceil} - p_{m-\lceil \frac{\varphi}{8}N \rceil}, p_m + p_{m+\lceil \frac{\eta}{2} \rceil} - p_{m-\lceil \frac{\varphi}{8}N \rceil}]$. Since this interval is of length $p_{m+\lceil \frac{\eta}{2} \rceil} - p_{m-\lceil \frac{\eta}{2} \rceil}$, it is weakly larger than E_η, and since preferences are *Non-Clumped*, X contains the ideal policy of at least one agent. If $p_{m+\lceil \frac{\varphi}{8}N \rceil} \in X$, then let $j = m + \lceil \frac{\varphi}{8}N \rceil$. If $p_{m+\lceil \frac{\varphi}{8}N \rceil} \notin X$, then let j be the left-most agent with an ideal policy in X. Since $p_{m+\lceil \frac{\varphi}{8}N \rceil} - p_m < p_m - p_{m-\lceil \frac{\varphi}{8}N \rceil}$, it follows that if $p_{m+\lceil \frac{\varphi}{8}N \rceil} \notin X$, then $p_j > p_{m+\lceil \frac{\varphi}{8}N \rceil}$. In either case, $p_j \geq p_{m+\lceil \frac{\varphi}{8}N \rceil}$ and

$$\widehat{p}_{ij} \in \left[\frac{p_m + p_{m-\lceil \frac{\eta}{2} \rceil}}{2}, \frac{p_m + p_{m+\lceil \frac{\eta}{2} \rceil}}{2} \right] \subset \left(p_{m-\lceil \frac{\eta}{2} \rceil}, p_{m+\lceil \frac{\eta}{2} \rceil} \right),$$

and thus, the difference in support for the two candidates is no more than η. Then i and j, by definition of η, want to run against each other.

Now we show that no other agent would want to enter the race.

With a support strategy σ^* that assigns no support to a third entrant, a candidate d with $p_d < p_i$ would get at most $m - \lceil \frac{\varphi}{8}N \rceil - 1$ votes if all citizens disregard σ^* and vote sincerely, whereas j would get no less than $m - \lceil \frac{\eta}{2} \rceil + 1$. The ratio of their respective support $\frac{S_d}{S_j}$ is no larger than

$$\frac{\frac{(1-\frac{\varphi}{4})}{2}N}{\frac{N-\eta}{2}} = \frac{(4-\varphi)N}{4(N-\eta)} = \frac{1-\frac{1}{4}\varphi}{1-\frac{\eta}{N}}.$$

Consider a sequence of societies of increasing size and the corresponding sequence of pairs of candidates i_N and j_N and third entrants d_N, where the subindex now identifies them as members of the society of size N. As $N \to \infty$, η_N goes to infinity, but $\frac{\eta_N}{N}$ converges to zero. Therefore, the upper bound on the ratio of the share of supporters of d_N over the share of supporters of j_N converges to $1 - \frac{1}{4}\varphi < 1$. Given a fixed ratio of support between two candidates, the probability that the candidate with the lesser share of support wins converges to zero. So a candidate d_N to the left of i_N will not win if N is sufficiently large. Then, $c > c^*$ guarantees that sure losers do not run.

Similarly for a candidate that tried to outflank j instead.

Now, a centrist candidate between i_N and j_N will collect at most the support of all citizens between p_{i_N} and p_{j_N}, that is, $2\lceil \frac{\varphi}{8}N \rceil$ if $j_N = p_{m_N+\lceil \frac{\varphi}{8}N \rceil}$ or $\lceil \frac{\varphi}{8}N \rceil$ and all the votes of citizens with ideal policies in $(p_{m_N}, p_{m_N} + p_{m_N - \lceil \frac{\eta N}{2} \rceil} - p_{m_N - \lceil \frac{\varphi}{8} N \rceil}]$ otherwise. But

$$(p_{m_N}, p_{m_N} + p_{m_N - \lceil \frac{\eta N}{2} \rceil} - p_{m_N - \lceil \frac{\varphi}{8} N \rceil}] \subset [p_{m_N}, 2p_{m_N} - p_{m_N - \lceil \frac{\varphi}{8} N \rceil}],$$

and since preferences are $\varphi - balanced$,

$$2p_{m_N} - p_{m_N - \lceil \frac{\varphi}{8} N \rceil} = p_{m_N} + (p_{m_N} - p_{m_N - \lceil \frac{\varphi}{8} N \rceil}) < p_{m_N} + p_{N+1-\lceil \frac{3\varphi}{8} N \rceil} - p_{m_N} = p_{N+1-\lceil \frac{3\varphi}{8} N \rceil}.$$

Thus, there are no more than $\frac{N+1}{2} - \lceil \frac{3}{8}\varphi N \rceil$ agents with an ideal policy in $(p_{m_N}, p_{m_N} + p_{m_N - \lceil \frac{\eta N}{2} \rceil} - p_{m_N - \lceil \frac{\varphi}{8} N \rceil}]$. Then, the centrist entrant gathers a total number of supports which is less than $\frac{N+1}{2} - 2\lceil \frac{\varphi}{8} N \rceil$. Since i gathers at least $\frac{N}{2} - \lceil \frac{\varphi}{8} N \rceil$ supporters, their ratio of support is no larger than $\frac{1-\frac{\varphi}{2}}{1-\frac{\varphi}{4}} = \frac{4-2\varphi}{4-\varphi} < 1$, so the probability that the centrist entrant wins converges to zero with the size of the electorate. For a sufficiently big society, $c > c^*$ guarantees that the entrant does not run.

For an even N, the proof is almost identical, with $i = m_l - \lceil \frac{\varphi}{8} N \rceil$,

$$X = \left[2p_{m_h - \lceil \frac{\eta}{2} \rceil} - p_{m_l - \lceil \frac{\varphi}{8} N \rceil}, 2p_{m_l + \lceil \frac{\eta}{2} \rceil} - p_{m_l - \lceil \frac{\varphi}{8} N \rceil}\right]$$

and j equal to $m_h + \lceil \frac{\varphi}{8} N \rceil$ if $p_{m_h + \lceil \frac{\varphi}{8} N \rceil} \in X$, and j equal to the left-most citizen with an ideal policy in X otherwise. ■

2.6.8 Proof of Lemma 9

Proof. By contradiction. Suppose i is a spoiler and i is the left-most candidate and let the second left-most candidate be j. If every citizen chooses support sincerely and i drops out of the race, all the support for i switches to support for j. The probability that i wins the election is now zero, the probability that j wins increases, and the probability that any of the other candidates wins is reduced. Since $v_i(p_j) \geq v_i(p_k)$ for any candidate $k \notin \{i, j\}$, it follows $\sum_{k \in C \setminus i} v_i(p_k) \frac{\Pr[W_k=1|C]}{\Pr[W_i=0|C]} \leq$

$\sum_{h\in C\backslash i} v_i(p_h)\Pr[W_h = 1|C\backslash i]$.

Since i is a spoiler,

$$\{b + v_i(p_i) - \sum_{k\in C\backslash i} v_i(p_k)\Pr[W_k = 1|C\backslash i]\}\Pr[W_i = 1|C] < c.$$

Thus:

$$\begin{aligned}\Pr[W_i &= 1|C]\{b + v_i(p_i) - \sum_{k\in C\backslash i} v_i(p_k)\Pr[W_k = 1|C\backslash i]\} \\ +\Pr[W_i &= 0|C]\{\sum_{k\in C\backslash i} v_i(p_k)\frac{\Pr[W_k = 1|C]}{\Pr[W_i = 0|C]} - \sum_{k\in C\backslash i} v_i(p_k)\Pr[W_k = 1|C\backslash i]\} < c,\end{aligned}$$

which, rearranging terms, implies

$$b\Pr[W_i = 1|C] + \sum_{k\in C} v_i(p_k)\{\Pr[W_k = 1|C] - \Pr[W_k = 1|C\backslash i]\} < c.$$

Therefore, i is better off not running as a candidate. ■

2.6.9 Proof of Proposition 11

Proof. For two candidates: If $i, j \in C$ have different support, it must be they have different ideal policies. Then, given Euclidean preferences and given that $p_i, p_j \in Q$ but $\frac{p_i+p_j}{2} \notin Q$, every citizen supports either i or j (no mixed support strategies). Suppose $\frac{S_i}{N} - \frac{S_j}{N} \geq \epsilon$. Then $S_i \geq \frac{1+\epsilon}{2}N$ and $S_j \leq \frac{1-\epsilon}{2}N$. Note that V_k follows a binomial distribution $Bi[S_k, 1-\mu]$ for $k = i, j$. By the Weak Law of Large Numbers, $\frac{V_i}{S_i}$ and $\frac{V_j}{S_j}$ converge to $1-\mu$. Therefore, $\frac{V_i}{V_j}$ converges to $\frac{S_i}{S_j}$, which is by assumption more than one. Therefore, with probability converging to one as N approaches infinity, i receives more valid votes and wins the election. For a sufficiently low probability of victory, j prefers not to run as candidate.

For three candidates: Let $C = \{i, j, h\}$, where $E[S_i] \geq E[S_j] \geq E[S_h]$.

Suppose $\frac{E[S_i]}{N} - \frac{E[S_h]}{N} \geq \epsilon$. By the Central Limit Theorem, the distribution of $\frac{V_i}{N}$ converges to the point $(1-\mu)\frac{E[S_i]}{N}$, similarly $\frac{V_h}{N}$ converges to $(1-\mu)\frac{E[S_h]}{N}$ and the probability that h wins converges to zero. Thus for a large enough N, h has to be a spoiler. By lemma 9, it must then be that $p_h \in (p_i, p_j)$ where both i and j are competitive candidates. In order for j to be competitive, j must win with some positive probability. Again by the CLT, the distribution of $\frac{V_j}{N}$ converges to the point $(1-\mu)\frac{E[S_j]}{N}$. Thus, in order for j to win with enough probability, $\frac{E[S_i]}{N} - \frac{E[S_j]}{N}$ must converge to zero as N approaches infinity. I next show that the probability that given that two candidates tie for victory one of the two is h converges to zero. Let δ be an arbitrarily small, positive number such that $\delta < \frac{\epsilon}{2}$. Let $\Delta = [1-\delta, 1+\delta]$.

The probability that h ties for victory or wins by one vote is:

$$\begin{aligned}\Pr[\frac{V_h}{V_i} &\in \Delta] \Pr[V_h - V_i \in \{0,1\} | \frac{V_h}{V_i} \in \Delta] \Pr[V_h > V_j | V_h - V_i \in \{0,1\}] \\ + \Pr[\frac{V_h}{V_j} &\in \Delta] \Pr[V_h - V_j \in \{0,1\} | \frac{V_h}{V_j} \in \Delta] \Pr[V_h > V_i | V_h - V_j \in \{0,1\}].\end{aligned}$$

The probability that i and j tie for victory is:

$$\Pr[\frac{V_j}{V_i} \in \Delta] \Pr[V_j = V_i | \frac{V_j}{V_i} \in \Delta] \Pr[V_j > V_h | V_j = V_h].$$

Let $\Lambda_1 = \Pr[\frac{V_h}{V_i} \in \Delta] \Pr[V_h > V_j | V_h - V_i \in \{0,1\}]$, let $\Lambda_2 = \Pr[\frac{V_h}{V_j} \in \Delta] \Pr[V_h > V_i | V_h - V_j \in \{0,1\}]$, let $\Lambda_3 = \Pr[\frac{V_j}{V_i} \in \Delta] \Pr[V_j > V_h | V_j = V_h]$. Using this notational shortcut, the ratio of the probability that h ties for victory over the probability that i and j tie for victory is:

$$\frac{\Lambda_1 \Pr[V_h - V_i \in \{0,1\} | \frac{V_h}{V_i} \in \Delta] + \Lambda_2 \Pr[V_h - V_j \in \{0,1\} | \frac{V_h}{V_j} \in \Delta]}{\Lambda_3 \Pr[V_j = V_i | \frac{V_j}{V_i} \in \Delta]}.$$

Since Λ_1 and Λ_2 converge to zero and Λ_3 converges to one as N approaches infinity, and

$$\frac{\Pr[V_h - V_i \in \{0,1\} | \frac{V_h}{V_i} \in \Delta]}{\Pr[V_j = V_i | \frac{V_j}{V_i} \in \Delta]} \text{ and } \frac{\Pr[V_h - V_j \in \{0,1\} | \frac{V_h}{V_j} \in \Delta]}{\Pr[V_j = V_i | \frac{V_j}{V_i} \in \Delta]}$$

are bounded, the previous ratio converges to zero. Therefore, given that a single vote is decisive, the probability that it is decisive to determine whether i or j wins converges to one, and for a large enough N, supporting h is wasted support. If all citizens choose support strategically, only citizens whose ideal policy is at a very small distance from $\widehat{p}_{ij}$ will support h. Let δ be such that any citizen with an ideal policy at a distance more than δ from $\widehat{p}_{ij}$ would strategically support i or j, not h. As N approaches infinity, this distance δ converges to zero, and the ideal policy of an strategic h supporter has to converge to $\widehat{p}_{ij}$. Let q' and q" be the two points in Q closest to $\frac{p_i+p_j}{2}$ such that $q' < \frac{p_i+p_j}{2} < q$". If N is large enough, $[\widehat{p}_{ij} - \delta, \widehat{p}_{ij} + \delta] \in (q', q")$ and no citizen, not even h herself will support h. Then, h cannot win and since the strategy profile is consistent, nor can h affect the election, and h prefers to drop out than to run with no supporters.

For more than three candidates: First note that there cannot be an equilibrium in which two or more candidates have a common ideal policy and one other candidate does not, because if so strategic citizens would concentrate their support in only one of the candidates with the same policy, and the others would drop out if they had no supporters. Second, suppose that there are two competitive candidates with distinct ideal policies. If the electorate is sufficiently large, any candidate trailing to the strongest candidate by more than ϵ in share of support has a probability of victory sufficiently close to zero to be a spoiler, whose ideal policy must be between that of two competitive candidates (lemma 9). The probability that a single vote for the spoiler affects the election given that a single vote is pivotal converges to zero. Therefore, by the same arguments as with three candidates, only citizens who are indifferent or close to indifferent about the two competitive candidates would support the spoiler. If the electorate is sufficiently large, there are no citizens with an ideal policy so close to the midpoint of the ideal policies of the competitive candidates, and then the spoiler who receives no votes drops out of the race. Suppose there are three or more competitive candidates and a weak candidate trailing by more than ϵ in support share. With Euclidean policies, no citizen can be simultaneously indifferent about three distinct policies. If the size of the electorate is large enough, the probability that a vote for the weakest candidate is decisive given that a vote is decisive is arbitrarily close to zero. Then, no citizen will support the weakest candidate, and this candidate will drop out, again by the inability to affect the election with no support derived from the assumption of consistency in the support strategy profile. ■

Chapter 3

United We Vote

ABSTRACT: This chapter studies the advantages that a coalition of agents obtain by forming a voting bloc to pool their votes and cast them all together. I identify the necessary and sufficient conditions for an agent to benefit from the formation of the voting bloc, both if the agent is a member of the bloc and if the agent is not part of the bloc. I also determine whether individual agents prefer to participate in or step out of the bloc, and I find the different optimal internal voting rules that aggregate preferences within the coalition.[1]

3.1 Introduction

Despite the advantages of collaboration, alliances are often broken, groups are dissolved, coalitions split, or they fail to be formed in the first place. Any union of heterogeneous agents may fail to act for the benefit of some of its members. Individual freedom of action is partially curtailed by joining a group and committing to follow its rules. This creates an incentive to abandon the group and proceed alone in a different course of action. There is a trade-off between the potential gains of group action and the sacrifice of individual freedom involved in group formation.

In this chapter I examine this trade-off in the context of political competition between agents

[1] This chapter is based on the material in Eguia [23].

who can communicate to form a coalition. I model a set of agents who face a vote over a choice of alternatives. I assume that a specific subset of the whole electorate can coalesce to coordinate the voting behavior of its members. Agents coalesce to increase the probability that their preferred outcome is chosen. For example, each country has a vote in the UN General Assembly. The European Union's 25 members could decide to coordinate their foreign policies, agreeing on a common voting position before UN meetings.

A crucial problem is how to choose the common position. Each country knows what it wants, and it also knows that a coalition of countries will have a better chance of getting what it asks for than a single country, hence there is an incentive to form such coalition. But if its members have conflicting preferences, what will the coalition stand for? If the coalition intends to act as a "voting bloc" and cast all its votes together, it requires an internal decision-making rule to aggregate the preferences of its members. This internal decision-making rule will map the possibly disparate preferences of the members into a single alternative for which all members of the coalition will vote.

The internal rule that maximizes the aggregate utility for the coalition is simple majority. However, only under certain conditions every member of the coalition benefits from forming a voting bloc if simple majority is chosen as the internal aggregating rule.

If all members do not benefit from simple majority, then the coalition must find other rules to aggregate the preferences of its members. Constitutional design studies the rules that determine how to change voting rules within a society. In our case, I assume that the coalition of agents can form a voting bloc to coordinate their votes only by unanimous agreement of all members. Thus every member of the coalition must be made better off, otherwise the coalition will not be able to function as a voting bloc, because some members will block the project. For instance, any one country of the EU can veto a new EU treaty that intended to unify the foreign policy of its members.

I find that the sufficient conditions for a supermajority internal rule to make every member of the coalition better off are less stringent than those needed for simple majority to do so. I also find that an "opt-out" rule benefits every member in some cases when supermajorities do not. Overall, for a very large set of possible preference profiles there exists some rule that satisfies every member in the coalition.

Imagine a successful coalition that has found one such rule and functions as a voting bloc, casting all its votes together according to the outcome of its internal decision-making rule and making every member better off than if everyone voted individually according to their own preferences. I find that under certain conditions some members will still have an incentive to leave the coalition (if that is possible). The coalition needs more than a rule that benefits everyone to function as a voting bloc: It needs to solve the collective-action problem in which members prefer others but not themselves to participate, although everyone is better off if all of them participate than if no one does.

Several non-cooperative theories of coalition formation with economic applications are surveyed in Carraro [14]. Closer to the motivation of this chapter, Buchanan and Tullock [12] analyze the costs and benefits of forming a coalition and praise the virtues of unanimity as internal voting rule. Barberá and Jackson [6] let agents choose among several rules and they define "self-stable" voting rules as those that will not be beaten by any other rule if the given voting rule is used to choose among rules. Maggi and Morelli's [39] study "self-enforcing" rules to determine whether collective action will be taken by a group of agents, and they conclude that no other rule but simple majority or unanimity is ever optimal.

A different approach to coalition formation comes from the voting power literature. Felsenthal and Machover [27], Gelman [28] and Feix, Lepelley, Merlin and Rouet [26] among others, analyze the probability of casting a decisive vote in an election and the effect of coalitions and alliances of voters over this probability. I focus instead on the probability of getting the desired outcome in the election. Laruelle and Valenciano [36] provide a rigorous analysis of the relation and differences between maximization of the probability of being pivotal, which is the object of study of the voting power literature, and maximization of the probability of success or satisfaction with the outcome, which is the approach I take. In essence, in the voting power literature the agents seek to maximize their probability of being able to alter the outcome; whereas, in my model the agents want to maximize the probability that the policy outcome coincides with their policy preference. The voting power approach applies to an scenario in which the goal of a voter is to win by one vote. If instead the goal of an agent is to win, be it by one vote or by a landslide, then my model applies.

I analyze the potential benefits of forming a voting bloc, coalescing with other agents to cast all votes in the same direction to make it more likely that a preferred alternative wins the election. I am interested in the effect of a voting bloc on the degree of satisfaction of its members, how the heterogeneity of such members may affect their gain in utility and which internal voting rules in the coalition may make the voting bloc satisfactory for a broader range of parameters.

These theoretical questions are particularly relevant to the ongoing debates about the need or desire for a common foreign policy in the EU, a purpose that was first vaguely stated in the Maastricht Treaty (1992) [2], but that has been recently the subject of much deeper debates and controversy during the negotiations towards a constitution (started in 2002) and will probably continue to be in the European political agenda for years to come. Therefore, I frequently refer to the EU as a motivating example along the exposition.

After introducing the model and showing that there is a surplus to be gained by forming a voting bloc in section 3.2, in section 3.3 I ask whether the formation of the voting bloc will benefit every member of the coalition. In section 3.4 I study an "opt-out" rule that allows one agent to stay out of the voting bloc and discuss under what conditions introducing such a rule will benefit all the members of the coalition. In section 3.5 I summarize the findings of the chapter. An appendix contains the proofs for the results in Sections 3.3 and 3.4.

3.2 The Model: Gains from Forming a Voting Bloc

Let $\mathcal{N} = \{1, 2, .., N\}$ be an assembly of voters, where $N \geq 5$ is finite and odd. The members of the assembly (legislators, countries, etc.) face a binary decision: either to keep the status quo, or to

[2] The Treaty on European Union, signed in Maastricht in 1992.

vote for an alternative a to replace it. All agents are called to vote either for a (*yes*), or against a (*no*). If the number of favorable votes is equal or higher than a threshold T, then a is implemented.

Each agent strictly prefers either the status quo or the alternative a, and I assume no intensity in preferences. Preferences over lotteries will simply be determined in favor of the lottery that assigns the higher weight to the preferred alternative.

The agents in assembly $\mathcal{N}$ are divided into two sets. A set of agents M with size N_M lack coordination powers and vote individually. The remaining agents can coordinate among themselves and may at wish form a voting bloc. I call this second set of agents the coalition C, which has size $N_C = N - N_M$, where N_C is odd.[3]

If each of its members agrees, coalition C forms a voting bloc. In this case coalition C will hold an internal meeting to predetermine its voting behavior in the general vote. In the internal meeting, all members of the coalition will vote *yes* or *no* according to their preferences for or against a. Then:

1. If the majority in this internal vote has strictly more than $r_C N_C$ votes, where $r_C \in [\frac{1}{2}, \frac{N_C - 1}{N_C}]$, then the majority prevails and all members of coalition C will vote as a bloc in the general assembly casting N_C votes according to the preferences (either *yes* or *no*) of the majority of the coalition. The outcome of the coalitional internal meeting is binding.

2. If the majority gathers no more than $r_C N_C$ votes in the internal vote, then the coalition fails to act as a bloc in the general assembly and all members are free to vote according to their individual preferences.

Note that threshold r_C defines the $r_C - majority$ rule used by the coalition to decide whether or not it will act as a bloc rolling its internal minorities. A threshold $r_C \in [\frac{1}{2}, \frac{1}{2} + \frac{1}{N_C})$ corresponds to simple majority, $r_C \in [\frac{1}{2} + \frac{1}{N_C}, \frac{N_C - 1}{N_C})$ to a supermajority and $r_C = \frac{N_C - 1}{N_C}$ to unanimity.[4]

Forming a voting bloc with unanimity as internal voting rule is in essence identical to not forming a voting bloc, because the coalition will only cast its votes as a bloc if all its members share the

[3] The assumption that N_C is odd is for technical convenience. Qualitative results are the same if N_C is even.
[4] I use the notation $x \in [a, b]$ to mean $a \leq x \leq b$ and $x \in (a, b)$ to mean $a < x < b$.

same preference, in which case all votes will be cast as they would in the absence of a voting bloc.

If the coalition does not form a voting bloc, then all the members of the coalition will vote according to their individual preferences in the general assembly.

Coalition members decide to form a voting bloc with an $r_C - majority$ internal voting rule before the alternative a is specified, so agents do not know if they will prefer alternative a or the status quo. Agents have no power over the specification of alternative a, which is exogenous.

Every agent i has a type t_i, which is the probability that agent i will prefer alternative a to the status quo, once alternative a is revealed.[5] This type can be interpreted as a propensity for change, or as a displeasure with the status quo in general. Let t_{-i} denote the vector of types of all agents in the society other than i. Types are common knowledge, and so are true preferences once agents learn what the alternative a is.

Each realization of preferences is independent from the others. Once alternative a is revealed, each of possibly many agents with a type t_0 has an independent probability t_0 of supporting alternative a. Typically several of them will end up supporting a, whereas some others will prefer the status quo.

If the coalition forms a voting bloc, in the internal vote, voting will be sincere and there will be no abstention. With simple majority as the internal decision making rule, if the number of *yes* votes surpasses the number of negative ones, then the whole coalition (now a voting bloc) will cast a total of N_C *yes* votes in the general vote which includes all agents in the society. If the number of *no* votes surpasses the number of favorable ones, then the coalition accordingly votes as a bloc casting N_C *no* votes in the general vote.

The voting bloc behavior I have described consists of rolling internal minorities to present a common front in the general vote, strengthening the position of the coalition's majority with the minority votes which are "converted" or "swayed" to the majoritarian camp, increasing the chances

[5] This probabilistic model of voter uncertainty was first considered by Rae [47] and developed by Badger [5] and Curtis [16].

of eventually getting the outcome the majority wishes (of course, in doing so the probability of getting what the minority wishes decreases).

Proposition 12 *Let type $t_i \in (0,1)$ for each agent i in the society. Then, for any $N_C \geq 3$ (number of agents in the coalition), $N_M \geq 2$ (number of agents not in the coalition), and T (threshold to accept alternative a), a coalition of N_C members strictly increases the sum of expected utilities of its members by forming a voting bloc with either simple majority or any supermajority as the internal voting rule. Simple majority rule is the internal voting rule that maximizes sum of expected utilities of the members of the voting bloc.*

Forming a voting bloc only has an effect in utilities if the formation of a bloc and the subsequent rolling of minority votes within the coalition alters the outcome of the general vote. If so, every member of the coalition who is in the coalitional majority benefits from the voting bloc formation, at the cost of every voter in the minority. Since the majority is by definition bigger than the minority, there are more members benefiting than suffering from the bloc, and since the intensity of preferences is set to be equal for every member, in the aggregate forming a voting bloc generates a surplus of utility for the coalition. Any other rule that in some cases fails to roll a minority is giving away this net gain in utility and therefore underperforms in comparison to simple majority in terms of aggregated gains in utility.[6]

It follows from proposition 12 that if all the members of the coalition share a common type, then forming a voting bloc increases the utility of every member in the coalition. Therefore, a homogeneous coalition of agents who have the same type should always form a voting bloc with simple majority as the internal voting rule to maximize their probability of winning the final vote in a larger electorate. Also from proposition 12, I derive the following corollary:

Corollary 13 *If all but one of the members of the coalition share a common type, all the homoge-*

[6]Beyond the sketched intuition provided here, the interested reader can find a formal proof of Proposition 12 and Proposition 14 below in the working paper version, Eguia [21], and a generalization of Proposition 12 to multiple coalitions in Proposition 32 in Chapter 4. These proofs follow in part the insight due to Curtis [16].

neous members benefit from the formation of a voting bloc.

Suppose all members in C except for i share a common type. Then if member i is in the rolled minority, more of the homogeneous members are in the majority benefiting from the rolling of votes the voting bloc imposes than in the hurt minority, thus in the aggregate the homogeneous members strictly benefit from the bloc. If member i is in the majority, there are at least the same number of homogeneous members in the majority as in the minority. Thus, in the aggregate the bloc is at worst neutral to the homogeneous members. Since both cases are possible, overall there is a surplus for the homogeneous members (maybe not so for the heterogeneous one).

In Sections 3.3 and 3.4 I will show under which conditions will every member of an heterogeneous coalition benefit from the formation of a voting bloc. I now ask whether the formation of a voting bloc benefits or harms the interests of the agents who are not part of the coalition. The answer will depend on the voting rule in the general election.

Proposition 14 *Let type $t_i \in (0,1)$ for each agent i in the society and let coalition C form a voting bloc with any internal rule other than unanimity. Then, if the voting rule in the general election is unanimity, every agent not in the coalition strictly benefits from the formation of the voting bloc. If the general voting rule is simple majority, there is a loss in aggregate utility for the agents not in the coalition and the society as a whole.*

If the rule in the general election is unanimity, then each agent has a veto power over changes to the status quo. If a coalition forms a voting bloc, it removes the veto power from its members, but not from non-members, who therefore benefit from the formation of a voting bloc by the coalition. If the rule in the general election is simple majority, a voting bloc will only change the outcome to make a minority of the society win. That is contrary to the interests of a majority of non-members of the coalition. As shown by Curtis [16], simple majority maximizes the social welfare of the society. Once a coalition forms a voting bloc, the wishes of the majority not always prevail, hence social welfare is reduced. From proposition 12 we know that if a coalition forms a voting bloc, the sum of utilities of its members increases; it must then be that the sum of utilities of non-members decreases with the formation of the voting bloc.

Even if the formation of a voting bloc is in the aggregate hurting non-members of the coalition, this effect will in general not be uniform. Some agents not in the coalition will win, whereas some lose expected utility if the coalition forms a voting bloc. For instance, suppose that the members of the coalition have types such that almost always the *yes* wins in the coalitional internal vote and a small but significant *no* minority is rolled. Then the bloc behavior by the coalition tilts the general vote in favor of alternative a. Agents with a high type, who are likely to prefer alternative a, will be then happy to see the coalition form a voting bloc.

To summarize, forming a voting bloc is inconsequential if the coalition uses unanimity as internal voting rule, but with any other rule, forming a bloc gives a surplus in utility to the coalition and simple majority is the internal voting rule that maximizes such surplus. Every agent in the rest of the society benefits from the formation of a bloc by the coalition if the general voting rule is unanimity, but if the general voting rule is simple majority these agents suffer an aggregate loss in utility.

In the next section I investigate under what conditions the coalition can reach unanimous agreement among its members to proceed with the formation of a voting bloc and appropriate the surplus in utility that comes with the voting bloc.

3.3 Achieving Consensus to Form a Voting Bloc

We wish to find an internal rule for the coalition to aggregate preferences in such a way that maximizes the aggregate utility of its members relative to a default in which the coalition uses unanimity as internal voting rule and all members always vote according to their true preferences in the general election. This internal rule must be such that in expectation every member prefers it to unanimity. Because each member of the coalition can block deviations from unanimity as internal voting rule, the coalition needs full consensus to use any other rule to aggregate its preferences.

Given an internal voting rule v, I say that v is *beneficial for* C if in expectation every member in C is weakly better off using v rather than unanimity as the internal voting rule and some member in C is strictly better off.

In short, a rule is beneficial for the coalition if it benefits all its members to adopt it instead of unanimity.

In this section I assume that the general election rule is simple majority and I investigate which $r_C - majority$ internal rules are beneficial for C. If there are several beneficial majority rules, I focus on whichever one maximizes the overall surplus for the coalition. Recall from proposition 12 that the internal voting rule that maximizes the aggregated utility for the coalition is simple majority. However, in an heterogeneous coalition, some members may not benefit from pooling votes in a voting bloc with simple majority.

I label unanimity rule as $\varnothing$ as a reminder that using unanimity as internal voting rule is identical to not forming a voting bloc, or no members joining the voting bloc.

I label as r_C a $r_C - majority$ rule in which every member of the coalition participates in the voting bloc, and minorities of size strictly less than $r_C N_C$ are rolled to join the position of the majority of the coalition in the general vote. Simple majority, denoted Sm, refers to the special case in which $r_C = \frac{1}{2}$.

For any internal voting rule v, let $EU_i[v]$ denote the expected utility for agent i if the coalition forms a voting bloc with v as internal voting rule.

Before I present the results, I need to make some assumptions on the types of the agents:

Assumption 1 *The number of favorable votes cast by the N_M agents not in coalition C follows a symmetric distribution around $\frac{N_M}{2}$ with some positive probability of casting a quantity of favorable votes different than $\frac{N_M}{2}$.*

This condition significantly relaxes the standard assumption in the voting power literature that all agents have a common type of 0.5.[7] Instead, it suffices that the N_M agents can be paired in such way that for each pair (j, j'), $t_j + t_{j'} = 1$, with at least one pair of agents with types strictly between zero and one. Let $f(x)$ denote the probability that the number of favorable votes for alternative a cast by the N_M agents not in the coalition is exactly x, and let $F(x) = \sum_{k=0}^{x} f(k)$ be the distribution function of the number of favorable votes cast by these N_M agents.

I make a milder assumption on the types of the members of coalition C. Namely, I assume that coalition C "leans toward" accepting alternative a. Let $g_{-i}(x)$ denote the probability that x members of $C\backslash\{i\}$, the coalition without i, prefer alternative a. Then I require the following:

Assumption 2 *For all $k \in [1, \frac{N_C-1}{2}]$ and for all $i \in C$, $g_{-i}(\frac{N_C-1}{2} + k) > g_{-i}(\frac{N_C-1}{2} - k)$.*

Note that $g_{-l}(k) = \sum_{\substack{A\subseteq C\backslash\{l\} \\ |A|=k}} [\prod_{\substack{i\in A \\ j\in C\backslash(A\cup\{l\})}} t_i(1-t_j)]$.

Assumption 2 states that given any $N_C - 1$ members of the coalition and given any particular majority-minority split of votes in this subset of the coalition, it is more probable that this majority in the subset is for the *yes* side. A sufficient condition for this assumption to hold is that excluding any member, we can pair the rest in such a way that for each pair (i, i'), $t_i + t_{i'} \geq 1$, with strict inequality for at least one pair.

Let $g_{-ij}(x)$ denote the probability that exactly x members of $C\backslash\{i, j\}$, the coalition without i or j, prefer alternative a. Formally,

$$g_{-lh}(k) = \sum_{\substack{A\subseteq C\backslash\{l,h\} \\ |A|=k}} [\prod_{\substack{i\in A \\ j\in C\backslash(A\cup\{l,h\})}} t_i(1-t_j)].$$

Let $l \in C$ be the agent with the lowest type in the coalition, so that $t_l \leq t_i$ for all $i \in C$. From Assumption 2, it follows that for all $i \in C$, and for all $k \in [0, \frac{N_C-3}{2}]$,

$$g_{-li}(\frac{N_C-1}{2} + k) > g_{-li}(\frac{N_C-3}{2} - k).$$

[7] See, for instance, Felsenthal and Machover [27].

I prove this technical claim in an appendix at the end of the chapter.

With these two assumptions on the types of the agents and simple majority as the voting rule in the general election, a member of the coalition will like to form a voting bloc with a $r_C - majority$ rule as internal voting rule if her type is "high enough": If a given member would benefit from forming a voting bloc with a $r_C - majority$, then every other member with a higher type would benefit even further.

Lemma 15 *Let $l, h \in C$ such that $t_h \geq t_l$. Then $EU_h[r_C] - EU_h[\varnothing] \geq EU_l[r_C] - EU_l[\varnothing]$.*

By lemma 15 we can focus only on the member with the lowest type to see if she benefits from the formation of a voting bloc with a $r_C - majority$. If she does, then every member in the coalition benefits from forming a voting bloc with a $r_C - majority$ rule:

Proposition 16 *Let $l \in C$ be the member with the lowest type. Then a $r_C - majority$ rule is beneficial as an internal voting rule for coalition C if and only if $t_l > t^{r_C,\varnothing}(t_{-l})$.*

In the appendix to this chapter I find the exact expression of $t^{r_C,\varnothing}(t_{-l})$ as a function of the types of the agents and the threshold r_C.

Since the coalition "leans" towards accepting alternative a, the majority within the coalition will be in favor of alternative a more often than not, with the result that the negative votes will be rolled more often than the favorable ones. Therefore it becomes more likely that alternative a wins the general election. Member l only likes such voting behavior if her type is high "enough," where the exact meaning of "enough" is given by the threshold in the proposition.

I illustrate proposition 16 with the aid of Figure 3-1, for the specific case of simple majority as the internal voting rule.

To be able to plot the threshold $t^{Sm,\varnothing}(t_{-l})$ with respect to only one variable, in Figure 3-1 I assume that the distribution of votes by the agents not in coalition C follows a binomial $Bi(N_M, \frac{1}{2})$ and that all the members of coalition C except l share a common type t_C, and I set $N_M = 176$ and

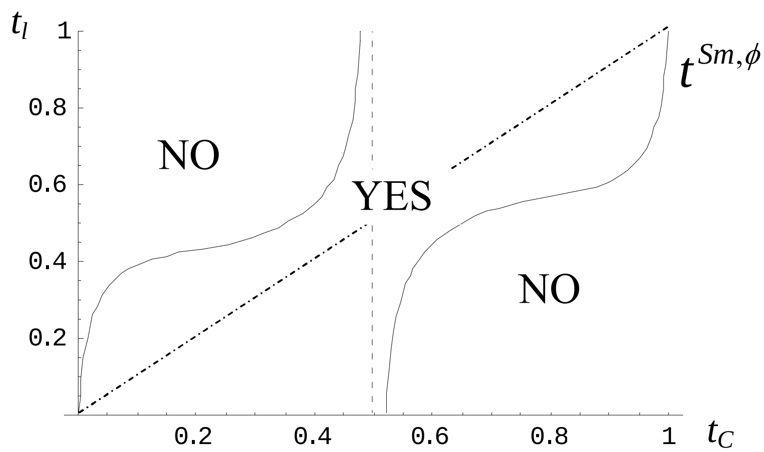

Figure 3-1: Consensus to form a voting bloc with simple majority.

$N_C = 25$ to approximate the European Union example. I use the same assumptions for all the other figures.

The model corresponds to the right half of the graph: if the common type t_C of the $N_C - 1$ members other than l is bigger than one half, member l supports the formation of a voting bloc with simple majority if the type t_l is above the depicted threshold. The left half of the picture is a symmetric case in which the coalition leans towards rejecting a. Then member l supports the formation of a bloc only if her type is below the threshold.

For some type profiles, simple majority is not *beneficial for C*. Nevertheless, the coalition can form a voting bloc with a supermajority internal voting rule that benefits every member. Using a more stringent supermajority internal voting rule enables the coalition to accommodate a more extreme member l with a low type t_l, making l benefit from the formation of a voting bloc with a $r_C - majority$ internal rule, even if l rejects a bloc with simple majority.

I depict this result in Figure 3-2, which shows four possible rules for a coalition the size of the EU: Simple majority, two-thirds majority, four-fifths majority, and nine-tenths majority.

Note how the range of parameters for which a voting bloc would benefit every member increases

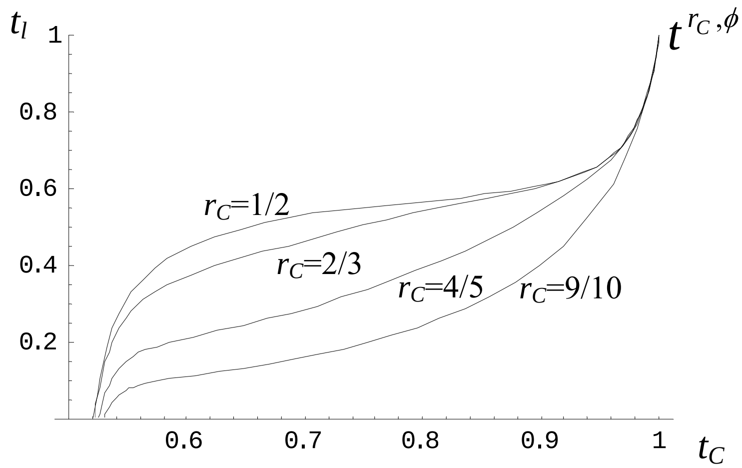

Figure 3-2: Supermajority rules.

as the supermajority rule becomes more stringent. However, since simple majority maximizes the overall surplus for the coalition, setting higher thresholds for approval of a common position diminishes the value of the voting bloc, although it may help to bring an outlier on board.

Aiming to maximize the utility of the coalition subject to not hurting any member, the optimizing solution is the lowest possible supermajority that would benefit (or at least leave indifferent) the member with the lowest type.

In the remainder of this section I investigate how changes in the size of the coalition or the heterogeneity of types of its members affect which rules the coalition will be able to use to the benefit of all its members.

I find that if the size of the coalition is too large, then no coalition in which all members but l share a common type $t_C > t_l$ can form a voting bloc with simple majority.

Proposition 17 *Let M be fixed. Let $t_l < t_C$ for $l \in C$ and $t_i = t_C$ for all $i \in C \backslash \{l\}$. There exists some $\overline{N}$ such that if $N_C > \overline{N}$, simple majority is not beneficial for C.*

As the size of the coalition becomes very large relative to N_M, the internal majority coincides with the external majority unless the coalition is almost evenly split. The coalition is more likely to vote for a than against a. Since member l is the member with the lowest type, conditional on the coalition being evenly split, member l is more likely to be against a, thus on the losing side. Therefore, if the coalition becomes so large that rolling its votes only affects the outcome when the coalition is almost evenly split, the member with the lowest type rejects the formation of a bloc with simple majority. In the limit, only a fully homogeneous coalition where every member has the same type could form a voting bloc with simple majority.

Beyond size, I ask how heterogeneity affects the chances of a coalition forming a voting bloc. We know that simple majority is beneficial for any homogeneous coalition, whereas heterogeneous coalitions may run into obstacles. Nevertheless, I show by means of an example that the possibility of forming a voting bloc with simple majority is not monotonic with heterogeneity.

Compare three coalitions with the same mean type and the same lowest type. I measure heterogeneity by the standard deviation of types. I find that the most homogeneous and the least homogeneous of the three coalitions cannot form a voting bloc with simple majority, whereas the intermediate one can.

Example 2 *Let all agents not in C have a type $t_M = 0.5$ and let there be 10 of them.*

Let C_1 be a coalition of agents with types {0.445, 0.75, 0.75, 0.75, 0.75}. The mean type is 0.689. The standard deviation 0.1368. If $C = C_1$, then $t^{Sm,\varnothing}(t_{-l}) = 0.4943$ and l rejects the formation of a voting bloc with simple majority as internal voting rule.

Let C_2 be another coalition of agents with types {0.445, 0.5, 0.5, 1, 1}, mean type 0.689, standard deviation 0.2847. If $C = C_2$, then $t^{Sm,\varnothing}(t_{-l}) = 0.441$ so forming a voting bloc with simple majority benefits every member of the coalition.

Let C_3 be yet another coalition of agents with types {0.445, 0.45, 0.55, 1, 1}, mean type 0.689, standard deviation 0.2869. If $C = C_3$, then $t^{Sm,\varnothing}(t_{-l}) = 0.446$ so once again member l vetoes the formation of a voting bloc with simple majority.

For coalitions C_1 and C_3 in Example 2, using a two-thirds majority or a 3/4 majority (or any other value of r_C that requires a majority of 4 to 1 to roll the minority) every member benefits from forming a voting bloc. In C_1, using $r_C = 3/4$, $t^{3/4,\varnothing}(t_{-l}) = 0.39$ so member l favors the formation of a voting bloc that rolls only minorities of size one. Similar results hold for coalition C_3.

Considering coalition C_1 in Example 2, I quantify the impact that the formation of a voting bloc would have over the outcome in the general election. I show the results in Table 3.1. The numbers represent the probability that the event indicated in each row occurs, given the internal rule the coalition uses. In the second column, the coalition uses unanimity or forms no bloc, in the third column it forms a bloc with simple majority and in the fourth column it uses a 3/4 majority.

Table 3.1: The impact of forming a voting bloc

	No bloc	1/2 maj	3/4 maj
a approved	69.52%	79.58%	73.45%
a approved given l likes a	79.82%	90.00%	84.72%
a approved given l dislikes a	61.26%	71.22%	64.42%
l satisfied with outcome	57.02%	56.02%	57.44%
$j \in C \backslash l$ satisfied with outcome	67.26%	74.70%	70.57%
$m \notin C$ satisfied with outcome	59.44%	53.51%	57.61%

Since the coalition leans towards a, forming a voting bloc makes approval of a more likely. All the members except for l benefit and all non-members are hurt forming a voting bloc. Member l is hurt if simple majority is used, so in order to benefit all its members, the coalition has to select a supermajority that makes l better off, but attenuates the advantage for all the other members. Forming a voting bloc with simple majority would have a substantial impact: the probability of approving a increases ten percentage points and the probability of getting the desired outcome out of the election would increase seven percentage points for all members of coalition C_1 but l. Using a 3/4 majority reduces this benefit of a voting bloc to roughly a half, but it makes all members of C_1 more likely to see their preference prevail in the general election.

In this section I have described the necessary and sufficient condition for a coalition to be able to form a voting bloc with a majority rule. I show that although simple majority always maximizes the aggregate surplus, there are type profiles for which simple majority is not beneficial for the coalition but some supermajority rules are and the coalition can choose one of them to gain some of the surplus of a voting bloc benefiting all its members.

3.4 An Opt-Out Rule

In this section I explore a more nuanced rule, which consists of forming a voting bloc with all but one of the members of the coalition. The excluded member does not participate in the internal vote of the voting bloc, but votes directly and according to her true preferences in the general election.

This scheme differs from expelling one member from the coalition in a crucial detail: The exclusion is voluntary, the member who does not participate in the voting bloc agrees to the formation of the voting bloc without her, and hence she opts to be out, or "opts-out." The member who opts-out has to benefit from the formation of the voting bloc by the other members, otherwise she would rather veto the whole project and keep unanimity in place as the voting rule to aggregate votes in the coalition.

Denote by *Out* the "Opt-Out for l" rule in which the member l with the lowest type does not participate in the voting bloc which is formed by every other member of the coalition and simple majority is chosen as internal voting rule.

Throughout this section I assume that the general election rule is simple majority.

The first result on opt-out rules considers the conditions under which the member of the coalition who stays out of the voting bloc benefits from its formation.

Proposition 18 *The formation of a voting bloc with simple majority rule by every member of coalition C except l, benefits member l if and only if $t_l > t^{Out,\varnothing}(t_{-l})$.*

I provide the expression of $t^{Out,\varnothing}(t_{-l})$ in the appendix at the end of the chapter.

The threshold function $t^{Out,\varnothing}(t_{-l})$ is always positive given our assumptions on types; it is not increasing with respect to the type of all other members of the coalition and it does not always converge to one as the types of the members of the coalition do. This last feature guarantees that in

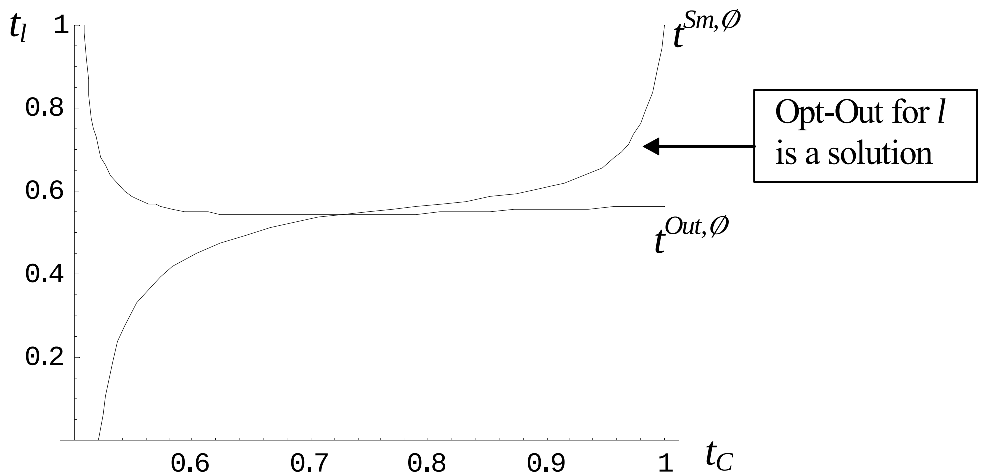

Figure 3-3: An opt-out rule as a solution.

some cases in which member l rejects forming a voting bloc with simple majority she benefits from the formation of a bloc with an "Opt-Out for l" rule. If all the other members also benefit from the "Opt-Out for l" rule, then this rule is *beneficial for C* and it offers a solution for a coalition which couldn't form a bloc with $r_C - majority$ rules. The next proposition states this result.

Proposition 19 *If* $5 \leq N_C \leq N_M + 1$, *there exist type profiles for which an "Opt-Out for l" internal voting rule is beneficial for* C *and no* $r_C - majority$ *rule is. If* $N = 3$ *or* $N_C > N_M + 1$, *there exists no type profile for which "Opt-Out for l" is beneficial for* C *and simple majority is not.*

If $N_C = 3$, allowing one member to step out reduces the bloc to size two, which is identical to not forming a bloc at all, or forming it with unanimity. If $N_C > N_M$, then the coalition acts as a dictator even if one member opts-out, thus the member who opts-out cannot be better off out of the voting bloc than in the voting bloc.

Figure 3-3 compares the threshold $t^{Out,\varnothing}(t_{-l})$ with $t^{Sm,\varnothing}(t_{-l})$ from the previous section. Notice that if types are in the area below the threshold $t^{Sm,\varnothing}(t_{-l})$ and above $t^{Out,\varnothing}(t_{-l})$, l would veto forming a voting bloc with simple majority if it included all the members. By allowing l to stay out, the coalition can form a voting bloc with simple majority with every other member and in expectation raise the utility of every member including l.

This result casts a favorable light over “opt-out” rules. On the other hand, “opt-out” rules have two setbacks: If the coalition is heterogeneous (and not just dichotomous with all members but l sharing a common type) and l opts out, it is possible that the member with the next lowest type opposes the formation of the reduced bloc. Allowing this member to opt-out as well may simply pass the problem to the next member until the bloc fully unravels and every member but the last two opt-out, which negates the purpose of a voting bloc.

Even if this unravelling does not take place, there is a second latent complication to opt-out rules: if the coalition allows for the member with the lowest type to opt-out, then other members may also request to opt-out, even if they benefit from the formation of a voting bloc, simply because they would benefit even more by opting-out. If the coalition lets every member join in or stay out of the voting bloc, then it faces a “free-rider” problem, where some members who would benefit from joining the voting bloc, may prefer to opt-out and passively take advantage of the pooling of votes by other coalition partners.

Definition 9 *Member l “free rides” if she would have benefitted from forming and participating in a voting bloc, but benefits even more as a result of opting-out.*

With no opt-out rules, there is no chance to free ride, since the coalition faces an “all-or-none” binary decision: either every member joins the voting bloc, or the bloc is not formed. If instead members can individually choose whether to join in or to stay out, some may choose not to join in. In the following Proposition I explore whether a member would prefer to participate in or to stay out of a voting bloc formed by every other member of the coalition.

Proposition 20 *Member l prefers to participate in a voting bloc formed by the coalition with simple majority as internal decision rule better than to opt-out and not participate in the pooling of votes by the rest of the members of the coalition if and only if $t_l > t^{Sm,Out}(t_{-l})$.*

I provide the exact expression of $t^{Sm,Out}(t_{-l})$ in the appendix to this chapter.

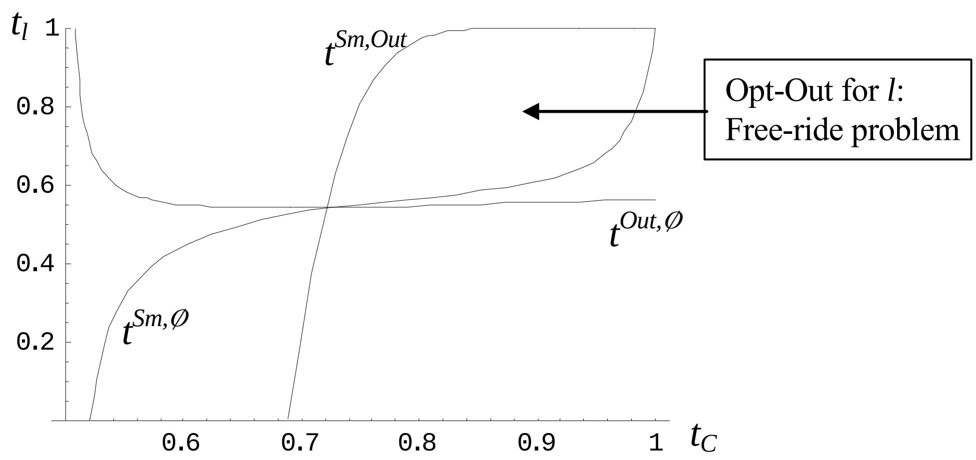

Figure 3-4: Opt-out and free ride.

From proposition 16 we obtained the condition for l to benefit from forming and participating in a voting bloc with simple majority. Proposition 20 now states when will member l prefer to opt-out from such a bloc. Combining Propositions 16 and 20 I obtain proposition 21, which demonstrates the downside of opt-out rules: They create a free-riding problem when member l would benefit from participating in the bloc but prefers to opt-out.

Proposition 21 *An "Opt-Out for l" rule creates a free-rider problem if and only if* $t^{Sm,\varnothing}(t_{-l}) < t_l < t^{Sm,Out}(t_{-l})$. *If* $5 \leq N_C \leq N_M + 1$, *there exist type profiles for which this condition is met. If* $N = 3$ *or* $N_C > N_M + 1$, *this condition cannot hold and free riding cannot occur.*

If $t_l < t^{Sm,Out}(t_{-l})$, member l prefers not to participate in the voting bloc and she free rides if participating would be better for l than not forming a voting bloc.

Let us visualize when a member will prefer to opt-out and free ride on her coalition partners with the aid of Figure 3-4. If type t_l is above $t^{Sm,\varnothing}(t_{-l})$ but below $t^{Sm,Out}(t_{-l})$, member l would have supported a voting bloc with simple majority and no opt-outs better than no bloc, but she prefers to opt-out if she can. If she opts-out, the overall utility for the coalition is reduced.

As a summary, allowing a member to opt-out can be a good solution in a coalition with great homogeneity of types and one outlier, but in many other instances it can generate free-rider problems, which are aggravated if the possibility to opt-out is extended to every member.

3.5 Conclusions

A coalition of agents who are a part of a larger electorate or assembly facing a vote may choose to form a voting bloc. The coalition will then pool its votes and cast them all together according to the outcome of an internal vote. I have shown that forming a voting bloc generates a surplus in the aggregate utility of the members of the coalition and I have checked that simple majority is the internal rule for the voting bloc that maximizes such surplus.

However, if there is heterogeneity among the members of the coalition, the surplus will not be evenly shared. In the absence of transfers, the formation of a voting bloc may be detrimental to some members of the coalition. Ordering members from the least likely to the most likely to support changes to the status quo, I find a single cutting point separating those members of the coalition who support the formation of a voting bloc, and those who reject it. This implies that either every agent in the coalition supports the voting bloc, or at most, agents at one tail of the distribution of types reject it.

Under the motivation that each member of the coalition can veto the formation of any kind of voting bloc, I have analyzed possible solutions to reach consensus to form a voting bloc with an alternative internal rule when some member of the coalition opposes a bloc with simple majority.

Using qualified majority rules (supermajorities) as the internal voting rules reduces the overall surplus in aggregate utility but it sometimes enables the coalition to achieve unanimous support for the formation of a voting bloc, when no such agreement was possible to form a bloc with simple majority. This result contrasts with the findings of Maggi and Morelli's model [39], in which only simple majority or unanimity are ever found to be optimal, but their focus is on homogeneous agents.

In the last section I have considered another solution: for some range of parameters, allowing an extreme agent who opposes the formation of a voting bloc to opt-out and not participate in the bloc is sufficient to achieve unanimous support (including support by the member who chooses to opt-out) for the formation of a voting bloc by the rest of the coalition.

These theoretical results have practical implications, suggesting that a collection of countries with some similarity in their policy preferences would do better by forging a common foreign policy that was not based in unanimity. In particular, each of the 25 members of the EU would be more likely to see its preference prevail at a UN Assembly meeting (or at any international forum that grants one vote per country) if the Union first predetermined how it will cast all its 25 votes according to an internal voting rule that rolled minorities within the EU.

A more interesting extension and generalization of this model to fit a wider set of applications consists on allowing several coalitions, not just one, to form voting blocs. Ideally, any subset of agents would be allowed to form a voting bloc and we would look for stable partitions of the space of agents into voting blocs. A very natural scenario in which voting blocs may occur is any legislature in which political parties may be formed. If we want to explain party formation, we need to allow for different parties to exist. Starting with a set of individual legislators, parties would be each one of the voting blocs that are formed. I pursue this approach in the next chapter of this thesis.

3.6 Appendix

3.6.1 Proof of Claim from Assumption 2

Assumption 2: For all $k \in [1, \frac{N_C-1}{2}]$ and for all $i \in C$, $g_{-i}(\frac{N_C-1}{2}+k) > g_{-i}(\frac{N_C-1}{2}-k)$.

We want to show that $g_{-li}(\frac{N_C-1}{2}+k) > g_{-li}(\frac{N_C-3}{2}-k)$ for any $i \in C$.

Let us first introduce some useful notation. For any set of agents $D \subseteq \mathcal{N}$, let $g^D(x)$ denote the probability that exactly x agents in D prefer alternative a, and let $g^D_{-i}(x)$ denote the probability that exactly x agents in $D\backslash\{i\}$ prefer alternative a. We then proceed in two steps.

Claim 22 *Let $D \subseteq \mathcal{N}$ be such that N_D is even and $g^D(\frac{N_D}{2} - k) \leq g^D(\frac{N_D}{2} + k)$ for any positive integer k. Let $l \in D$ be such that $t_l \leq t_i$ for any $i \in D$. Let $D' = l' \cup D \backslash l$ with $t_{l'} > t_l$. Then*

$$g^{D'}(\frac{N_D}{2} - k) \leq g^{D'}(\frac{N_D}{2} + k) \textit{ for any positive integer } k.$$

Proof. Consider $D\backslash l$ and let G^D_{-l} be the distribution function of the number of agents in $D\backslash\{l\}$ who prefer alternative a. Its probability mass function g^D_{-l} is determined by the aggregation of independent Bernoulli experiments, hence it is unimodal, as shown by Darroch [17]. Let y denote the mode. Since preferences are independent, for any number x and any agent l,

$$g^D(x) = t_l g^D_{-l}(x-1) + (1-t_l)g^D_{-l}(x).$$

Hence

$$g^{D\prime}(x) - g^D(x) = (t_{l'} - t_l)[g^D_{-l}(x-1) - g^D_{-l}(x)]. \tag{3.1}$$

Since g^D_{-l} is unimodal, for any $x > y$ it follows that $g^D_{-l}(x) \leq g^D_{-l}(x-1)$ and expression (3.1) is positive. For any $x \leq y$ it follows that $g^D_{-l}(x) \geq g^D_{-l}(x-1)$ and expression (3.1) is negative. It also follows from the unimodality of g^D_{-l} that the modes of g^D and $g^{D'}$ are either y or $y+1$, call them y^D and $y^{D'}$ respectively. Further, since $g^D(\frac{N_D}{2} - 1) \leq g^D(\frac{N_D}{2} + 1)$, it must be that $\frac{N_D}{2} \leq y^D \leq y+1$. Hence $\frac{N_D}{2} - 1 \leq y$ and $g^{D'}(\frac{N_D}{2} - k) \leq g^D(\frac{N_D}{2} - k)$ for any positive integer k.

For k such that $\frac{N_D}{2} + k > y^{D'} \geq y$, it follows that

$$g^{D'}(\frac{N_D}{2} - k) \leq g^D(\frac{N_D}{2} - k) \leq g^D(\frac{N_D}{2} + k) \leq g^{D'}(\frac{N_D}{2} + k),$$

where the first and third inequalities hold by the sign of expression (3.1) and the second inequality by assumption.

For k such that $\frac{N_D}{2} + k \leq y^{D'}$, $g^{D'}(\frac{N_D}{2} - k) \leq g^{D'}(\frac{N_D}{2} + k)$ by the unimodality of $g^{D'}$. ■

The proof of this first step is similar for D of odd size, with $g^{D'}(\frac{N_D-1}{2}-k) \leq g^{D'}(\frac{N_D+1}{2}+k)$ for any non negative integer k. The second step follows.

Claim 23 *Let $D \subseteq \mathcal{N}$ be such that N_D is even and let $t_l \leq t_i$ for any $i \in D$. Suppose that $g^D(\frac{N_D}{2}-k) \leq g^D(\frac{N_D}{2}+k)$ for all positive k. Then $g^D_{-l}(\frac{N_D}{2}-1-k) \leq g^D_{-l}(\frac{N_D}{2}+k)$ for any non-negative integer k.*

Proof. Note that the statement is immediately true if $t_l \geq 1/2$. We only need to prove it for $t_l < 1/2$. First construct $D' = \{D \cup l'\} \backslash l$ with $t_{l'} = 1/2$. By Claim 22, for any non-negative integer k, coalition D' satisfies

$$g^{D'}(\frac{N_D}{2}-k) \leq g^{D'}(\frac{N_D}{2}+k). \tag{3.2}$$

The rest of the proof proceeds by induction. First, for $k = \frac{N_D}{2}-1$, I prove that $g^{D'}_{-l'}(0) \leq g^{D'}_{-l'}(N_D-1)$.

From inequality (3.2), $g^{D'}(0) \leq g^{D'}(N_D)$. It follows $(1-t_{l'})g^D_{-l}(0) \leq t_{l'}g^D_{-l}(N_D-1)$. Since $t_{l'} = 1/2$, then $g^D_{-l}(0) \leq g^D_{-l}(N_D-1)$.

Suppose that $g^D_{-l}(\frac{N_D-2}{2}-k) \leq g^D_{-l}(\frac{N_D}{2}+k)$ holds for $k = k'$. I then show that it holds for $k = k'-1$.

From $g^{D'}(\frac{N_D}{2}-k') \leq g^{D'}(\frac{N_D}{2}+k')$, get

$$t_{l'}g^{D'}_{-l'}(\frac{N_D-2}{2}-k')+(1-t_{l'})g^{D'}_{-l'}(\frac{N_D}{2}-k') \leq t_{l'}g^{D'}_{-l'}(\frac{N_D}{2}+k'-1)+(1-t_{l'})g^{D'}_{-l'}(\frac{N_D}{2}+k'), \tag{3.3}$$

$$g^{D'}_{-l'}(\frac{N_D-2}{2}-k')+g^{D'}_{-l'}(\frac{N_D}{2}-k') \leq g^{D'}_{-l'}(\frac{N_D}{2}+k'-1)+g^{D'}_{-l'}(\frac{N_D}{2}+k'), \tag{3.4}$$

$$g^{D'}_{-l'}(\frac{N_D}{2}-k') \leq g^{D'}_{-l'}(\frac{N_D}{2}+k'-1), \tag{3.5}$$

$$g^D_{-l}(\frac{N_D-2}{2}-(k'-1)) \leq g^D_{-l}(\frac{N_D}{2}+k'-1). \tag{3.6}$$

Expression (3.3) implies (3.4) because $t_{l'} = (1-t_{l'}) = 1/2$. Expression (3.4) implies (3.5) because $g^D_{-l}(\frac{N_D-2}{2}-k') \leq g^D_{-l}(\frac{N_D}{2}+k')$ by assumption. Expression (3.6) is merely a reformulation of (3.5).

The induction argument is then complete. ■

Finally, for an arbitrary $i \in C\backslash\{l\}$, let $D = C\backslash\{i\}$. Then by Claim 23, $g_{-li}(\frac{N_C-3}{2} - k) \leq g_{-li}(\frac{N_C-1}{2}+k)$ for any non-negative integer k. Assuming strict inequality of $g^D(\frac{N_D}{2}-k) \leq g^D(\frac{N_D}{2}+k)$ at the statement of Claim 22 yields the strict inequality

$$g_{-li}(\frac{N_C-3}{2} - k) < g_{-li}(\frac{N_C-1}{2} + k)$$

at the end result of Claim 23, as desired. We can now prove lemma 15.

3.6.2 Proof of Lemma 15

Let A and B be two internal voting rules for coalition C. Let us define two functions, which depend on the rules A and B, and the vector of types t_{-i}:

$\alpha_i(B, A, t_{-i})$ = Probability that, given that member i prefers *yes,* the outcome in the general election is *yes* if the coalition uses rule B and *no* if it uses rule A.

$\beta_i(B, A, t_{-i})$ = Probability that, given that member i prefers *no,* the outcome in the general election is *no* if the coalition uses rule B and *yes* if it uses rule A.

I will use these functions in most of the proofs in this Appendix. In lemma 15, we want to show that $EU_h[r_C] - EU_h[\varnothing] - (EU_l[r_C] - EU_l[\varnothing]) \geq 0$.

Proof. $EU_i[r_C] - EU_i[\varnothing] =$

$$t_i\alpha_i(r_C, \varnothing, t_{-i}) + (1-t_i)\beta_i(r_C, \varnothing, t_{-i}) - t_i\alpha_i(\varnothing, r_C, t_{-i}) - (1-t_i)\beta_i(\varnothing, r_C, t_{-i}) \text{ for } i \in \{l, h\}.$$

In Step 1 I show that

$$t_h\alpha_h(r_C, \varnothing, t_{-h}) + (1-t_h)\beta_h(r_C, \varnothing, t_{-h}) - t_l\alpha_l(r_C, \varnothing, t_{-l}) - (1-t_l)\beta_l(r_C, \varnothing, t_{-l})$$

is positive. In Step 2, I show that

$$-t_h\alpha_h(\varnothing, r_C, t_{-h}) - (1-t_h)\beta_h(\varnothing, r_C, t_{-h}) + t_l\alpha_l(\varnothing, r_C, t_{-l}) + (1-t_l)\beta_l(\varnothing, r_C, t_{-l})$$

is also positive, thus adding all the terms, $EU_h[r_C] - EU_h[\varnothing] - (EU_l[r_C] - EU_l[\varnothing])$ is also positive.

Step 1:

Let $\lceil x \rceil$ denote the smallest integer equal or larger than x and similarly let $\lfloor x \rfloor$ denote the largest integer smaller or equal to x. With this convention, for $i = \{l, h\}$,

$$\alpha_i(r_C, \varnothing, t_{-i}) = \sum_{k=\lfloor r_C N_C \rfloor}^{N_C-2} g_{-i}(k)[F(\frac{N_M + N_C - 3}{2} - k) - F(\frac{N_M - N_C - 1}{2})].$$

Then, writing $g_{-l}(k)$ as $g_{-l}(k) = t_h g_{-lh}(k-1) + (1-t_h)g_{-lh}(k)$, I obtain:

$$t_l\alpha_l(r_C, \varnothing, t_{-l}) =$$
$$= \sum_{k=\lfloor r_C N_C \rfloor}^{N_C-2} t_l[t_h g_{-lh}(k-1) + (1-t_h)g_{-lh}(k)][F(\frac{N_M + N_C - 3}{2} - k) - F(\frac{N_M - N_C - 1}{2})],$$

and

$$t_h\alpha_h(r_C, \varnothing, t_{-h}) =$$
$$= \sum_{k=\lfloor r_C N_C \rfloor}^{N_C-2} t_h[t_l g_{-lh}(k-1) + (1-t_l)g_{-lh}(k)][F(\frac{N_M + N_C - 3}{2} - k) - F(\frac{N_M - N_C - 1}{2})],$$

so:

$$t_h\alpha_h(r_C, \varnothing, t_{-h}) - t_l\alpha_l(r_C, \varnothing, t_{-l}) =$$
$$= \sum_{k=\lfloor r_C N_C \rfloor}^{N_C-2} [(t_h - t_l)g_{-lh}(k)][F(\frac{N_M + N_C - 3}{2} - k) - F(\frac{N_M - N_C - 1}{2})].$$

which relabeling the counter in the summation becomes:

$$\sum_{k=\lfloor r_C N_C \rfloor - \frac{N_C-1}{2}}^{\frac{N_C-3}{2}} [(t_h - t_l) g_{-lh}(\frac{N_C-1}{2}+k)][F(\frac{N_M}{2}-k-1) - F(\frac{N_M - N_C - 1}{2})].$$

Now, noting that for $i = \{l, h\}$,

$$\beta_i(r_C, \varnothing, t_{-i}) = \sum_{k=1}^{\lceil (1-r_C) N_C \rceil - 1} g_{-i}(k)[F(\frac{N_M + N_C - 1}{2}) - F(\frac{N_M + N_C - 1}{2} - k)],$$

and omitting a very similar step I directly obtain that

$$(1-t_h)\beta_h(r_C, \varnothing, t_{-h}) - (1-t_l)\beta_l(r_C, \varnothing, t_{-l})$$
$$= - \sum_{k=1}^{\lceil (1-r_C) N_C \rceil - 1} [(t_h - t_l) g_{-lh}(k-1)][F(\frac{N_M + N_C - 1}{2}) - F(\frac{N_M + N_C - 1}{2} - k)],$$

which, since $\lceil r_C N_C \rceil + \lfloor (1-r_C) N_C \rfloor = N_C$, relabeling the counter in the summation becomes:

$$(1-t_h)\beta_h(r_C, \varnothing, t_{-h}) - (1-t_l)\beta_l(r_C, \varnothing, t_{-l})$$
$$= - \sum_{k=\lfloor r_C N_C \rfloor - \frac{N_C-1}{2}}^{\frac{N_C-3}{2}} [(t_h - t_l) g_{-lh}(\frac{N_C-3}{2} - k)][F(\frac{N_M + N_C - 1}{2}) - F(\frac{N_M}{2} + k)].$$

By Assumption 1,

$$F(\frac{N_M + N_C - 1}{2}) - F(\frac{N_M}{2} + k) = F(\frac{N_M}{2} - k - 1) - F(\frac{N_M - N_C - 1}{2}),$$

it follows

$$(1-t_h)\beta_h(r_C, \varnothing, t_{-h}) + t_h \alpha_h(r_C, \varnothing, t_{-h}) - (1-t_l)\beta_l(r_C, \varnothing, t_{-l}) - t_l \alpha_l(r_C, \varnothing, t_{-l})$$
$$= \sum_{k=\lfloor r_C N_C \rfloor - \frac{N_C-1}{2}}^{\frac{N_C-3}{2}} [(t_h - t_l)[F(\frac{N_M + N_C - 1}{2}) - F(\frac{N_M}{2} + k)][g_{-lh}(\frac{N_C-1}{2} + k) - g_{-lh}(\frac{N_C-3}{2} - k)],$$

which is positive by Assumption 2 as shown by Claim 23.

Step 2:

Noting that for $i = \{l, h\}$,

$$\alpha_i(\varnothing, r_C, t_{-i}) = \sum_{k=0}^{\lceil (1-r_C)N_C \rceil - 2} g_{-i}(k)[F(\frac{N_M + N_C - 1}{2}) - F(\frac{N_M + N_C - 3}{2} - k)] \text{ and}$$

$$\beta_i(\varnothing, r_C, t_{-i}) = \sum_{k=\lfloor r_C N_C \rfloor + 1}^{N_C - 1} g_{-i}(k)[F(\frac{N_M + N_C - 1}{2} - k) - F(\frac{N_M - N_C - 1}{2})],$$

and repeating the same steps as in Step 1, I get:

$$t_h \alpha_h(\varnothing, r_C, t_{-h}) - t_l \alpha_l(\varnothing, r_C, t_{-l})$$
$$= \sum_{k=0}^{\lceil (1-r_C)N_C \rceil - 2} [(t_h - t_l) g_{-lh}(k)][F(\frac{N_M + N_C - 1}{2}) - F(\frac{N_M + N_C - 3}{2} - k)],$$

which relabeling the counter in the summation becomes:

$$t_h \alpha_h(\varnothing, r_C, t_{-h}) - t_l \alpha_l(\varnothing, r_C, t_{-l})$$
$$= \sum_{k=\lfloor r_C N_C \rfloor - \frac{N_C - 1}{2}}^{\frac{N_C - 3}{2}} [(t_h - t_l) g_{-lh}(\frac{N_C - 3}{2} - k)][F(\frac{N_M + N_C - 1}{2}) - F(\frac{N_M}{2} + k)],$$

and

$$(1 - t_h)\beta_h(\varnothing, r_C, t_{-h}) - (1 - t_l)\beta_l(\varnothing, r_C, t_{-l}) =$$
$$= -\sum_{k=\lfloor r_C N_C \rfloor + 1}^{N_C - 1} [(t_h - t_l) g_{-lh}(k - 1)][F(\frac{N_M + N_C - 1}{2} - k) - F(\frac{N_M - N_C - 1}{2})],$$

which, relabeling once again, becomes

$$(1-t_h)\beta_h(\varnothing, r_C, t_{-h}) - (1-t_l)\beta_l(\varnothing, r_C, t_{-l})$$

$$= -\sum_{k=\lfloor r_C N_C \rfloor - \frac{N_C-1}{2}}^{\frac{N_C-3}{2}} [(t_h - t_l)g_{-lh}(\frac{N_C-1}{2}+k)][F(\frac{N_M}{2}-k-1) - F(\frac{N_M-N_C-1}{2})].$$

Therefore,

$$-(1-t_h)\beta_h(\varnothing, r_C, t_{-h}) - t_h\alpha_h(\varnothing, r_C, t_{-h}) + (1-t_l)\beta_l(\varnothing, r_C, t_{-l}) + t_l\alpha_l(\varnothing, r_C, t_{-l}) =$$

$$\sum_{k=\lfloor r_C N_C \rfloor - \frac{N_C-1}{2}}^{\frac{N_C-3}{2}} [(t_h - t_l)[F(\frac{N_M}{2}-k-1) - F(\frac{N_M-N_C-1}{2})][g_{-lh}(\frac{N_C-1}{2}+k) - g_{-lh}(\frac{N_C-3}{2}-k)]$$

which is also positive by Assumption 2.

It follows that $EU_h[r_C] - EU_h[\varnothing] - (EU_l[r_C] - EU_l[\varnothing]) \geq 0$. ■

3.6.3 Proof of Lemma 24

To shorten the proofs of Propositions 16, 18 and 20 I introduce a second lemma:

Lemma 24 *Given two internal voting rules A and B for coalition C, a member $i \in C$ is indifferent between A and B if:*

$t_i = \frac{\beta_i(A,B,t_{-i})-\beta_i(B,A,t_{-i})}{\alpha_i(B,A,t_{-i})-\alpha_i(A,B,t_{-i})+\beta_i(A,B,t_{-i})-\beta_i(B,A,t_{-i})}$.

Proof. $EU_i(B) - EU_i(A) = t_i[\alpha_i(B,A,t_{-i}) - \alpha_i(A,B,t_{-i})] + (1-t_i)[\beta_i(B,A,t_{-i}) - \beta_i(A,B,t_{-i})]$, where the functions α_i and β_i are as defined in the proof of lemma 15.

Equating to zero and solving for t_i I get:

$t_i = \frac{\beta_i(A,B,t_{-i})-\beta_i(B,A,t_{-i})}{\alpha_i(B,A,t_{-i})-\alpha_i(A,B,t_{-i})+\beta_i(A,B,t_{-i})-\beta_i(B,A,t_{-i})}$. ■

3.6.4 Proof of Proposition 16

Let simple majority be the general election rule. Then the formation of a voting bloc by the coalition with r_C as the internal decision rule benefits member $l \in C$ if and only if $t_l > t^{r_C,\varnothing}(t_{-l})$.

Proof. By lemma 24, member l is indifferent between a voting bloc with a $r_C - majority$ or a voting bloc with unanimity (identical to no voting bloc) if:

$t^{r_C,\varnothing}(t_{-l}) = \frac{\beta_l(\varnothing,r_C,t_{-l})-\beta_l(r_C,\varnothing,t_{-l})}{\alpha_l(r_C,\varnothing,t_{-l})-a_l(\varnothing,r_C,t_{-l})+\beta_l(\varnothing,r_C,t_{-l})-\beta_l(r_C,\varnothing,t_{-l})}$,

where:

$$\begin{aligned}
\alpha_l(r_C,\varnothing,t_{-l}) &= \sum_{k=\lfloor r_C N_C \rfloor}^{N_C-2} g_{-l}(k)[F(\frac{N_M+N_C-3}{2}-k)-F(\frac{N_M-N_C-1}{2})]; \\
\beta_l(r_C,\varnothing,t_{-l}) &= \sum_{k=1}^{\lceil(1-r_C)N_C\rceil-1} g_{-l}(k)[F(\frac{N_M+N_C-1}{2})-F(\frac{N_M+N_C-1}{2}-k)]; \\
a_l(\varnothing,r_C,t_{-l}) &= \sum_{k=0}^{\lceil(1-r_C)N_C\rceil-2} g_{-l}(k)[F(\frac{N_M+N_C-1}{2})-F(\frac{N_M+N_C-3}{2}-k)]; \\
\text{and } \beta_l(\varnothing,r_C,t_{-l}) &= \sum_{k=\lfloor r_C N_C \rfloor+1}^{N_C-1} g_{-l}(k)[F(\frac{N_M+N_C-1}{2}-k)-F(\frac{N_M-N_C-1}{2})].
\end{aligned}$$

The derivative with respect to t_l of the surplus for member l generated by the voting bloc with internal voting rule r_C is equal to the denominator of $t^{r_C,\varnothing}(t_{-l})$, which relabeling the counter in the four summations in the denominator, is equal to:

$$\begin{aligned}
& \sum_{k=\lfloor r_C N_C \rfloor-\frac{N_C-1}{2}}^{\frac{N_C-3}{2}} g_{-l}(\frac{N_C-1}{2}+k)[F(\frac{N_M}{2}-k-1)-F(\frac{N_M-N_C-1}{2})] \\
- & \sum_{k=\lfloor r_C N_C \rfloor-\frac{N_C-1}{2}}^{\frac{N_C-3}{2}} g_{-l}(\frac{N_C-1}{2}-k)[F(\frac{N_M+N_C-1}{2})-F(\frac{N_M}{2}+k)] \\
+ & \sum_{k=\lfloor r_C N_C \rfloor-\frac{N_C-3}{2}}^{\frac{N_C-1}{2}} g_{-l}(\frac{N_C-1}{2}+k)[F(\frac{N_M}{2}-k)-F(\frac{N_M-N_C-1}{2})] \\
- & \sum_{k=\lfloor r_C N_C \rfloor-\frac{N_C-3}{2}}^{\frac{N_C-1}{2}} g_{-l}(\frac{N_C-1}{2}-k)[F(\frac{N_M+N_C-1}{2})-F(\frac{N_M}{2}+k-1)]
\end{aligned}$$

$$= \sum_{k=\lfloor r_C N_C \rfloor - \frac{N_C-3}{2}}^{\frac{N_C-1}{2}} g_{-l}(\frac{N_C-1}{2}+k)[F(\frac{N_M}{2}-k) - 2F(\frac{N_M-N_C-1}{2}) + F(\frac{N_M}{2}-k-1)]$$
$$+g_{-l}(\lfloor r_C N_C \rfloor)[F(\frac{N_M+N_C-3}{2} - \lfloor r_C N_C \rfloor) - F(\frac{N_M-N_C-1}{2})$$
$$- \sum_{k=\lfloor r_C N_C) \rfloor - \frac{N_C-3}{2}}^{\frac{N_C-1}{2}} g_{-l}(\frac{N_C-1}{2}-k)[2F(\frac{N_M+N_C-1}{2}) - F(\frac{N_M}{2}+k-1) - F(\frac{N_M}{2}+k)]$$
$$-g_{-l}(N - \lfloor r_C N_C \rfloor)[F(\frac{N_M+N_C-1}{2}) - F(\frac{N_M-N_C+1}{2} + \lfloor r_C N_C \rfloor)].$$

Then note that, by assumption,

$$F(\frac{N_M}{2}-k-1) = 1 - F(\frac{N_M}{2}+k); F(\frac{N_M}{2}-k) = 1 - F(\frac{N_M}{2}+k-1),$$

and $2F(\frac{N_M-N_C-1}{2}) = 2 - 2F(\frac{N_M+N_C-1}{2})$.

Substitute accordingly to get:

$$\sum_{k=\lfloor r_C N_C \rfloor - \frac{N_C-3}{2}}^{\frac{N_C-1}{2}} \left\{ \begin{array}{c} [g_{-l}(\frac{N_C-1}{2}+k) - g_{-l}(\frac{N_C-1}{2}-k)] \\ {[2F(\frac{N_M+N_C-1}{2}) - F(\frac{N_M-1}{2}+k) - F(\frac{N_M}{2}+k-1)]} \end{array} \right\}$$
$$+g_{-l}(\lfloor r_C N_C \rfloor) \sum_{k=0}^{N_C-2-\lceil r_C(N_C-1)\rceil} f(\frac{N_M-N_C-1}{2}+k)$$
$$-g_{-l}(N_C-1-\lfloor r_C N_C \rfloor) \sum_{k=\lfloor r_C N_C \rfloor+1}^{N_C-1} f(\frac{N_M-N_C+1}{2}+k) =$$

$$\sum_{k=\lfloor r_C N_C \rfloor - \frac{N_C-3}{2}}^{\frac{N_C-1}{2}} \left\{ \begin{array}{c} [g_{-l}(\frac{N_C-1}{2}+k) - g_{-l}(\frac{N_C-1}{2}-k)] \\ {[2F(\frac{N_M+N_C-1}{2}) - F(\frac{N_M-1}{2}+k) - F(\frac{N_M}{2}+k-1)]} \end{array} \right\}$$
$$+[g_{-l}(\lfloor r_C N_C \rfloor) - g_{-l}(N_C-1-\lfloor r_C N_C \rfloor)] \sum_{k=0}^{N_C-2-\lfloor r_C N_C \rfloor} f(\frac{N_M-N_C+1}{2}+k).$$

Since $\lfloor r_C N_C \rfloor$ is bigger than $N_C - 1 - \lfloor r_C N_C \rfloor$ for $r_C \geq \frac{1}{2}$, if follows

$$g_{-l}(\lfloor r_C N_C \rfloor) \geq g_{-l}(N_C - \lfloor r_C N_C \rfloor),$$

and thus the denominator is positive.

Therefore,

$$EU_l[r_C] - EU_l[\varnothing] > 0 \Longleftrightarrow t_l > t^{r_C,\varnothing}.$$

Then, by lemma 15, $EU_l[r_C] - EU_l[\varnothing] > 0 \Longrightarrow EU_i[r_C] - EU_i[\varnothing] > 0$ for all $i \in C$.

As a corollary note, if the internal voting rule is simple majority, then $\lfloor r_C N_C \rfloor = \frac{N_C - 1}{2}$, and the threshold $t^{r_C,\varnothing}(t_{-l})$ becomes $t^{Sm,\varnothing}(t_{-l})$, which is equal to

$$\frac{\sum_{k=\frac{N_C+1}{2}}^{N_C-1} g_{-l}(k)[F(\frac{N_M+N_C-1}{2} - k) - F(\frac{N_M-N_C-1}{2})] - \sum_{k=1}^{\frac{N_C-1}{2}} g_{-l}(k)[F(\frac{N_M+N_C-1}{2}) - F(\frac{N_M+N_C-1}{2} - k)]}{\sum_{k=1}^{\frac{N_C-1}{2}} [g_{-l}(\frac{N_C-1}{2} + k) - g_{-l}(\frac{N_C-1}{2} - k)][2F(\frac{N_M+N_C-1}{2}) - F(\frac{N_M}{2} + k) - F(\frac{N_M}{2} + k - 1)]}.$$

■

3.6.5 Proof of Proposition 17

Proof. Let N_M be fixed. Let all the $N_C - 1$ members in $C\backslash\{l\}$ have a common type t_C. Then

$$EU_l[Sm] - EU_l[\varnothing] =$$

$$t_l \sum_{k=0}^{\frac{N_M}{2}-1} g_{-l}(\frac{N_C-1}{2}+k)F(\frac{N_M}{2}-k-1)$$
$$+(1-t_l)\sum_{k=0}^{\frac{N_M}{2}-1} g_{-l}(\frac{N_C-1}{2}-k)[1-F(\frac{N_M}{2}+k)]$$
$$-t_l\sum_{k=0}^{\frac{N_M}{2}-1} g_{-l}(\frac{N_C-3}{2}-k)[1-F(\frac{N_M}{2}+k)]$$
$$-(1-t_l)\sum_{k=0}^{\frac{N_M}{2}-1} g_{-l}(\frac{N_C+1}{2}+k)[F(\frac{N_M}{2}-k-1)].$$

Since $F(\frac{N_M}{2}-k-1) = [1-F(\frac{N_M}{2}+k)]$, this is equal to:

$$\sum_{k=0}^{\frac{N_M}{2}-1}\binom{N_C-1}{\frac{N_C-1}{2}-k}[t_l t_C^{\frac{N_C-1}{2}+k}(1-t_C)^{\frac{N_C-1}{2}-k} + (1-t_l)t_C^{\frac{N_C-1}{2}-k}(1-t_C)^{\frac{N_C-1}{2}+k}]F(\frac{N_M}{2}-k-1)$$
$$-\sum_{k=0}^{\frac{N_M}{2}-1}\binom{N_C-1}{\frac{N_C-3}{2}-k}[t_l t_C^{\frac{N_C-3}{2}-k}(1-t_C)^{\frac{N_C+1}{2}+k} + (1-t_l)t_C^{\frac{N_C+1}{2}+k}(1-t_C)^{\frac{N_C-3}{2}-k}]F(\frac{N_M}{2}-k-1),$$

where $\binom{a}{b} = \frac{a!}{b!(a-b)!}$.

Let t^* the value of t_l such that the expression above is equal to zero. Then:

$$\sum_{k=0}^{\frac{N_M}{2}-1} t^*\left\{\frac{1}{N_C-1-2k}[t_C^{2k+1}(1-t_C)-t_C(1-t_C)^{2k+1}] - \frac{1}{N_C+2k+1}[(1-t_C)^{2k+2}-t_C^{2k+2}]\right\}$$
$$= \sum_{k=0}^{\frac{N_M}{2}-1}\left(\frac{t_C^{2k+2}}{N_C+2k+1} - \frac{(1-t_C)^{2k+1}}{N_C-2k-1}\right).$$

Now break this equation into $\frac{N_M}{2}$ different equations and impose that for each $k \in \{0, \frac{M}{2} - 1\}$,

$$
\begin{aligned}
& t^* \left\{ \frac{1}{N_C - 2k - 1} [t_C^{2k+1}(1 - t_C) - t_C(1 - t_C)^{2k+1}] - \frac{1}{N_C + 2k + 1} [(1 - t_C)^{2k+2} - t_C^{2k+2}] \right\} \\
= \quad & \frac{t_C^{2k+2}}{N_C + 2k + 1} - \frac{(1 - t_C)^{2k+1}}{N_C - 2k - 1}
\end{aligned}
$$

A solution to this system of equations (with just one unknown) also solves the original equation. For each individual equation:

$$
t^* = \frac{(N_C - 2k - 1)t_C^{2k+2} - (N_C + 2k + 1)t_C(1 - t_C)^{2k+1}}{t_C(1 - t_C)[t_C^{2k} - (1 - t_C)^{2k}](N_C + 2k + 1) - (N_C - 2k - 1)[(1 - t_C)^{2k+2} - t_C^{2k+2}]},
$$

which, as $N_C \longrightarrow \infty$, converges to

$$
\frac{t_C^{2k+2} - t_C(1 - t_C)^{2k+1}}{t_C^{2k+1}(1 - t_C) - t_C(1 - t_C)^{2k+1} - (1 - t_C)^{2k+2} + t_C^{2k+2}} = t_C.
$$

So $\lim_{N_C -> \infty} t^{Sm,\varnothing}(t_{-l}) = t_C$. Since by assumption $t_l < t_C$, this implies that for N_C large enough, $t_l < t^{Sm,\varnothing}(t_{-l})$ and then member l would be hurt if C forms a voting bloc with simple majority. ■

3.6.6 Proof of Proposition 18

Proof. Let *Out* denote the internal voting rule for the coalition under which $N_C - 1$ members form a voting bloc with simple majority and member l stays out of the bloc and does not pool her vote with the rest of the coalition. By lemma 24 member l is indifferent between rules *Out* and $\varnothing$ if:

$$
t^{Out,\varnothing}(t_{-l}) = \frac{\beta_l(\varnothing, Out, t_{-l}) - \beta_l(Out, \varnothing, t_{-l})}{\alpha_l(Out, \varnothing, t_{-l}) - \alpha_l(\varnothing, Out, t_{-l}) + \beta_l(\varnothing, Out, t_{-l}) - \beta_l(Out, \varnothing, t_{-l})}, \text{ where:}
$$

$$\begin{aligned}
\alpha_l(Out, \varnothing, t_{-l}) &= \sum_{k=\frac{N_C+1}{2}}^{N_C-2} g_{-l}(k)[F(\frac{N_M+N_C-3}{2}-k)-F(\frac{N_M-N_C-1}{2})]; \\
\beta_l(Out, \varnothing, t_{-l}) &= \sum_{k=1}^{\frac{N_C-3}{2}} g_{-l}(k)[F(\frac{N_M+N_C-1}{2})-F(\frac{N_M+N_C-1}{2}-k)]; \\
\alpha_l(\varnothing, Out, t_{-l}) &= \sum_{k=1}^{\frac{N_C-3}{2}} g_{-l}(k)[F(\frac{N_M+N_C-3}{2})-F(\frac{N_M+N_C-3}{2}-k)]; \\
\text{and } \beta_l(\varnothing, Out, t_{-l}) &= \sum_{k=\frac{N_C+1}{2}}^{N_C-2} g_{-l}(k)[F(\frac{N_M+N_C-1}{2}-k)-F(\frac{N_M-N_C+1}{2})].
\end{aligned}$$

Relabeling the counters in the summations, the denominator

$$\alpha_l(Out, \varnothing, t_{-l}) - \alpha_l(\varnothing, Out, t_{-l}) + \beta_l(\varnothing, Out, t_{-l}) - \beta_l(Out, \varnothing, t_{-l})$$

is equal to:

$$\begin{aligned}
&\sum_{k=1}^{\frac{N_C-3}{2}} g_{-l}(\frac{N_C-1}{2}+k)[F(\frac{N_M}{2}-k-1)-F(\frac{N_M-N_C-1}{2})] \\
&-\sum_{k=1}^{\frac{N_C-3}{2}} g_{-l}(\frac{N_C-1}{2}-k)[F(\frac{N_M+N_C-3}{2})-F(\frac{N_M}{2}+k-1)] \\
&+\sum_{k=1}^{\frac{N_C-3}{2}} g_{-l}(\frac{N_C-1}{2}+k)[F(\frac{N_M}{2}-k)-F(\frac{N_M-N_C+1}{2})] \\
&-\sum_{k=1}^{\frac{N_C-3}{2}} g_{-l}(\frac{N_C-1}{2}-k)[F(\frac{N_M+N_C-1}{2})-F(\frac{N_M}{2}+k)].
\end{aligned}$$

Since

$$\begin{aligned}
&[F(\frac{N_M}{2}-k)-F(\frac{N_M-N_C+1}{2})+F(\frac{N_M}{2}-k-1)-F(\frac{N_M-N_C-1}{2})] = \\
&[F(\frac{N_M+N_C-1}{2})-F(\frac{N_M}{2}+k)+F(\frac{N_M+N_C-3}{2})-F(\frac{N_M}{2}+k-1)],
\end{aligned}$$

the denominator simplifies to

$$\sum_{k=1}^{\frac{N_C-3}{2}} \left\{ \begin{array}{c} [g_{-l}(\frac{N_C-1}{2}+k) - g_{-l}(\frac{N_C-1}{2}-k)] \\ [F(\frac{N_M}{2}-k) - F(\frac{N_M-N_C+1}{2}) + F(\frac{N_M}{2}-k-1) - F(\frac{N_M-N_C-1}{2})] \end{array} \right\},$$

and $t^{Out,\varnothing}(t_{-l})$ is equal to:

$$\frac{\sum_{k=1}^{\frac{N_C-3}{2}} g_{-l}(\frac{N_C-1}{2}+k)[F(\frac{N_M}{2}-k) - F(\frac{N_M-N_C+1}{2})] - g_{-l}(\frac{N_C-1}{2}-k)[F(\frac{N_M+N_C-1}{2}) - F(\frac{N_M}{2}+k)]}{\sum_{k=1}^{\frac{N_C-3}{2}} \left\{ \begin{array}{c} [g_{-l}(\frac{N_C-1}{2}+k) - g_{-l}(\frac{N_C-1}{2}-k)] \\ [F(\frac{N_M}{2}-k) - F(\frac{N_M-N_C+1}{2}) + F(\frac{N_M}{2}-k-1) - F(\frac{N_M-N_C-1}{2})] \end{array} \right\}}.$$

The difference in utility for agent l between the formation of a bloc without l and no bloc at all is

$$EU_l[Out] - EU_l[\varnothing] =$$

$$= t_l[\alpha_l(Out, \varnothing, t_{-l}) - \alpha_l(\varnothing, Out, t_{-l})] + (1 - t_l)[\beta_l(Out, \varnothing, t_{-l}) - \beta_l(\varnothing, Out, t_{-l})].$$

The derivative of $EU_l[Out] - EU_l[\varnothing]$ with respect to t_l coincides with the denominator of the threshold $t^{Out,\varnothing}(t_{-l})$, which is positive. Therefore, for t_l above the threshold $t^{Out,\varnothing}(t_{-l})$, member l prefers the formation of a voting bloc in which l does not participate better than not forming any bloc at all; whereas for t_l below $t^{Out,\varnothing}(t_{-l})$, member l prefers to form no bloc than to form a bloc in which l does not participate. ■

3.6.7 Proof of Proposition 19

Let V be the set of type profiles $(t_1, t_2, ..., t_N)$ satisfying Assumptions 1 and 2. Let $V^{r_C} \subset V$ be the subset of type profiles such that a $r_C - majority$ rule is beneficial for C, let $(V^{r_C})^C$ be its complement such that $V^{r_C} \cup (V^{r_C})^C \equiv V$ and let $V^{Out} \subset V$ be the subset of type profiles such that an "Opt-Out for l" rule is beneficial for C. Then, for any N_M and for any $5 \leq N_C \leq N_M + 1$,

$(V^{r_C})^C \cap V^{Out}$ is not empty.

Proof. Suppose $5 \le N_C \le N_M + 1$, $t_l = (1-\delta)$ and $t_j = (1-\varepsilon)\ \forall j \in C\backslash\{l\}$. Then:

$$EU_l[Out] - EU_l[\varnothing] =$$

$$\begin{aligned}
&= (1-\delta) \sum_{k=\lfloor r_C(N_C-1)\rfloor+1}^{N_C-2} g_{-l}(k)[F(\frac{N_M+N_C-3}{2}-k) - F(\frac{N_M-N_C-1}{2})] \\
&+\delta \sum_{k=1}^{\lceil(1-r_C)(N_C-1)\rceil-1} g_{-l}(k)[F(\frac{N_M+N_C-1}{2}) - F(\frac{N_M+N_C-1}{2}-k)] \\
&-(1-\delta) \sum_{k=1}^{\lceil(1-r_C)(N_C-1)\rceil-1} g_{-l}(k)[F(\frac{N_M+N_C-3}{2}) - F(\frac{N_M+N_C-3}{2}-k)] \\
&-\delta \sum_{k=\lfloor r_C(N_C-1)\rfloor+1}^{N_C-2} g_{-l}(k)[F(\frac{N_M+N_C-1}{2}-k) - F(\frac{N_M-N_C+1}{2})].
\end{aligned}$$

As ε converges to zero, $\frac{g_{-l}(k)}{g_{-l}(N_C-2)}$ converges to zero for any $k < N_C - 2$ and $EU_l[Out] - EU_l[\varnothing]$ converges to:

$(1-\delta)g_{-l}(N_C-2)[F(\frac{N_M-N_C+1}{2}) - F(\frac{N_M-N_C-1}{2})] - \delta g_l(N_C-2)[F(\frac{N_M-N_C+3}{2}) - F(\frac{N_M-N_C+1}{2})]$

which is positive for a sufficiently low δ, provided that $f(\frac{N_M-N_C+1}{2}) > 0$.

Then, there exist a $\delta > 0$ and $\overline{\varepsilon} > 0$ such that for all $\varepsilon < \overline{\varepsilon}$, $EU_l[Out] - EU_l[\varnothing] > 0$.

Therefore, if $5 \le N_C \le N_M + 1$, and the types of all the members but l converge to 1, member l with type $t_l = (1-\delta)$ benefits from a voting bloc with an "Opt-Out for l" rule. Given that all the other members of C share a common type, they all benefit from forming a voting bloc without l. Since $t^{r_C,\varnothing}(t_{-l})$ converges to 1, member l would not benefit from a $r_C - majority$ internal voting rule in a voting bloc that includes every member. It follows that if $\varepsilon < \overline{\varepsilon}$, any profile of types in which $t_l = (1-\delta)$ and $t_j = (1-\varepsilon)\ \forall j \in C\backslash\{l\}$ is in V^{Out} but not in V^{r_C}, thus $V^{Out} \not\subseteq V^{r_C}$.

If $N_C > N_M + 1$, then $EU_l[Sm] - EU_l[Out] = 2g_l(\frac{N_C-1}{2})[F(\frac{N_M+N_C-1}{2}) - F(\frac{N_M}{2})] > 0$, thus $V^{Out} \subset V^{Sm}$. If $N_C = 3$, then Out coincides with $\varnothing$ and V^{Out} is empty. ■

3.6.8 Proof of Proposition 20

Proof. By lemma 24, member l will be indifferent between participating in the voting bloc or opting out if

$$t^{Sm,Out}(t_{-l}) = \frac{\beta_l(Out, Sm, t_{-l}) - \beta_l(Sm, Out, t_{-l})}{\alpha_l(Sm, Out, t_{-l}) - \alpha_l(Out, Sm, t_{-l}) + \beta_l(Out, Sm, t_{-l}) - \beta_l(Sm, Out, t_{-l})},$$

where

$$\begin{aligned}
\alpha_l(Sm, Out, t_{-l}) &= g_{-l}(\frac{N_C - 1}{2})[F(\frac{N_M}{2} - 1) - F(\frac{N_M - N_C - 1}{2})], \\
\beta_l(Sm, Out, t_{-l}) &= g_{-l}(\frac{N_C - 1}{2})[F(\frac{N_M + N_C - 1}{2}) - F(\frac{N_M}{2})], \\
\alpha_l(Out, Sm, t_{-l}) &= \sum_{k=0}^{\frac{N_C-3}{2}} g_{-l}(k) f(\frac{N_M + N_C - 1}{2}), \text{ and} \\
\beta_l(Out, Sm, t_{-l}) &= \sum_{k=\frac{N_C+1}{2}}^{N_C-1} g_{-l}(k) f(\frac{N_M - N_C + 1}{2}).
\end{aligned}$$

Since $F(\frac{N_M}{2} - 1) - F(\frac{N_M - N_C - 1}{2}) = F(\frac{N_M + N_C - 1}{2}) - F(\frac{N_M}{2})$ and $f(\frac{N_M + N_C - 1}{2}) = f(\frac{N_M - N_C + 1}{2})$ it follows that $\alpha_l(Sm, Out, t_{-l}) = \beta_l(Sm, Out, t_{-l})$ and we can simplify the denominator to

$$\begin{aligned}
\alpha_l(Out, Sm, t_{-l}) - \beta_l(Out, Sm, t_{-l}) &= \left[\sum_{k=\frac{N_C+1}{2}}^{N_C-1} g_{-l}(k) - \sum_{k=0}^{\frac{N_C-3}{2}} g_{-l}(k) \right] f(\frac{N_M + N_C - 1}{2}) \\
&= \sum_{k=1}^{\frac{N_C-1}{2}} [g_{-l}(\frac{N_C - 1}{2} + k) - g_{-l}(\frac{N_C - 1}{2} - k)] f(\frac{N_M + N_C - 1}{2}),
\end{aligned}$$

and

$$t^{Sm,Out}(t_{-l}) = \frac{\sum_{k=\frac{N_C+1}{2}}^{N_C-1} g_{-}(k) f(\frac{N_M - N_C + 1}{2}) - g_{-l}(\frac{N_C - 1}{2})[F(\frac{N_M + N_C - 1}{2}) - F(\frac{N_M}{2})]}{\sum_{k=1}^{\frac{N_C-1}{2}} [g_{-l}(\frac{N_C - 1}{2} + k) - g_{-l}(\frac{N_C - 1}{2} - k)] f(\frac{N_M + N_C - 1}{2})}.$$

The advantage for member l of staying in is $EU_l[Sm] - EU_l[Out]$

$$= t_l[\alpha_l(Sm, Out, t_{-l}) - a_l(Out, Sm, t_{-l})] + (1 - t_l)[\beta_l(sm, Out, t_{-l}) - \beta_l(Out, Sm, t_{-l})].$$

Its derivative with respect to t_l coincides with the denominator of $t^{Sm,Out}(t_{-l})$. Since

$$g_{-l}(\frac{N_C - 1}{2} + k) > g_{-l}(\frac{N_C - 1}{2} - k) \text{ for all } k \in [1, \frac{N_C - 1}{2}],$$

the denominator and thus the derivative are positive. Therefore, member l prefers to stay in if type t_l is above the threshold $t^{Sm,Out}(t_{-l})$ and member l prefers to opt-out than to stay in if $t_l < t^{Sm,Out}(t_{-l})$. ■

3.6.9 Proof of Proposition 21

Proof. The first statement comes straightforward from Propositions 16 and 20. For the second one, suppose $t_i = t_C$ $\forall i \in C$. Then all members benefit from the formation of a voting bloc with simple majority: $EU_i[Sm] > EU_i[\varnothing]$ $\forall i \in C$. The extra gains of stepping out for member l when simple majority is the general voting rule are:

$$\begin{aligned} EU_l[Out] - EU_l[Sm] = t_C \sum_{k=0}^{\frac{N_C-3}{2}} g_{-l}(k) f(\frac{N_M + N_C - 1}{2}) + (1 - t_C) \sum_{k=\frac{N_C+1}{2}}^{N_C-1} g_{-l}(k) f(\frac{N_M - N_C + 1}{2}) \\ -p g_l(\frac{N_C - 1}{2})[F(\frac{N_M}{2} - 1) - F(\frac{N_M - N_C - 1}{2})] \\ -(1 - t_C) g_{-l}(\frac{N_C - 1}{2})[F(\frac{N_M + N_C - 1}{2}) - F(\frac{N_M}{2})]. \end{aligned}$$

For $t_C \in (1/2, 1)$ and any N_C, N_M :

$$
\begin{aligned}
&EU_l[Out] - EU_l[Sm] \\
&> (1-t_C)\sum_{k=\frac{N_C+1}{2}}^{N_C-1} g_{-l}(k) f(\frac{N_M - N_C + 1}{2}) - g_{-l}(\frac{N_C-1}{2})[F(\frac{N_M+N_C-1}{2}) - F(\frac{N_M}{2})] \\
&> (1-t_C) f(\frac{N_M - N_C + 1}{2}) g_{-l}(N_C - 1) - \frac{1}{2} g_{-l}(\frac{N_C-1}{2}),
\end{aligned}
$$

which letting $\alpha = f(\frac{N_M - N_C + 1}{2})$ and $\beta = \frac{(N_C-1)!}{\frac{(N_C-1)}{2}!\frac{(N_C-1)}{2}!}$ is equal to:

$$\alpha(1-t_C)t_C^{N_C-1} - \tfrac{1}{2}\beta p^{(N_C-1)/2}(1-t_C)^{(N_C-1)/2} = (1-t_C)t_C^{(N_C-1)/2}[\alpha p^{(N_C-1)/2} - \tfrac{1}{2}\beta(1-t_C)^{\frac{N_C-3}{2}}],$$

which is positive if and only if

$$2\alpha p^{\frac{N_C-1}{2}} \geq \beta(1-t_C)^{\frac{N_C-3}{2}} \iff \frac{t_C^{\frac{N_C-1}{2}}}{(1-t_C)^{\frac{N_C-3}{2}}} \geq \frac{\beta}{2\alpha} \iff \frac{t_C}{(1-t_C)^{\frac{N_C-3}{N_C-1}}} \geq \left(\frac{\beta}{2\alpha}\right)^{\frac{2}{N_C-1}}.$$

Letting $\gamma > 0$ be any number such that $t_C \geq \gamma$, this last inequality will be satisfied if

$$
\begin{aligned}
\frac{\gamma}{(1-t_C)^{\frac{N_C-3}{N_C-1}}} \quad &\geq \quad \left(\frac{\beta}{2\alpha}\right)^{\frac{2}{N_C-1}} \iff (1-t_C)^{\frac{N_C-3}{N_C-1}} \leq \gamma\left(\frac{2\alpha}{\beta}\right)^{\frac{2}{N_C-1}} \iff (1-t_C) \leq \gamma\left(\frac{2\alpha}{\beta}\right)^{\frac{2}{N_C-3}} \\
&\iff \quad t_C \geq 1 - \gamma\left(\frac{2\alpha}{\beta}\right)^{\frac{2}{N_C-3}},
\end{aligned}
$$

which is less than one for $N_C \in [5, M+1]$, provided that $f(\frac{N_M - N_C + 1}{2}) > 0$.

Note that if $N_C = 3$, then $\left(\frac{2\alpha}{\beta}\right)^{\frac{2}{N_C-3}} = 0$; whereas if $N_C > N_M + 1$, $\alpha = 0$.

If $N_C = 3$, then $EU_l[Out] = EU_l[\varnothing]$, thus $EU_l[Sm] > EU_l[\varnothing]$ implies $EU_l[Sm] > EU_l[Out]$.

If $N_C > N_M + 1$, then $f(\frac{N_M + N_C - 1}{2}) = f(\frac{N_M - N_C + 1}{2}) = 0$, thus

$$EU_l[Out] - EU_l[Sm] = -pg_l(\frac{N_C-1}{2})F(\frac{N_M}{2} - 1) - (1-t_C)g_{-l}(\frac{N_C-1}{2})[1 - F(\frac{N_M}{2})] < 0.$$

■

Chapter 4

Voting Blocs, Coalitions and Parties

ABSTRACT: In this chapter I study the strategic implications of coalition formation in an assembly. A coalition forms a voting bloc to coordinate the voting behavior of its members, acting as a single player and affecting the policy outcome. I prove that there exist stable endogenous voting bloc structures, and in an assembly with two parties I show how the incentives to join a bloc depend on the types of the agents, the sizes of the parties, and the rules the blocs use to aggregate preferences. I also provide an empirical application of the model to the US Supreme Court and I show that justices face a strategic incentive to coalesce into voting blocs.

4.1 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 coordinating their voting behavior and forming a voting bloc. 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 factions at faculty meetings in an academic department, to alliances of countries in international relations or 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.

However, agents are not identical and the benefits of forming a voting bloc are not equally shared by all. Some members of a voting bloc may prefer to leave the bloc, making it unstable. Who benefits when agents with diverse preferences form a voting bloc? What makes a voting bloc stable? What configuration of voting blocs do we expect to find in an assembly with heterogeneous voters? These are some of the questions that I address in this chapter, modeling an assembly with a finite number of agents who can coordinate with each other to form voting blocs before they vote to pass or reject a policy proposal.

My theory adds a novel insight about endogenous party formation. 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.

In the first part of the chapter I consider an assembly with two exogenously given parties, one on each side of the political spectrum, and I analyze whether or not every member of a party has an incentive to accept the party discipline depending on factors such as the types of the agents, the polarization of the assembly, the sizes of the parties, the internal rule that a party uses to aggregate the preferences of its members, and the process that leads to the formation of a voting bloc.

I find that in each party there is one extreme party member who is the least likely to benefit from the coordination of votes in her party, and this extreme agent determines whether or not the party can form a stable voting bloc with a given internal aggregation rule. I show that for some preference profiles a party cannot form a stable voting bloc that always imposes party discipline on its members, but it can form a stable voting bloc with laxer party discipline using an internal voting rule that lets members vote freely when there is substantial disagreement within the party. I also show that for some other preference profiles, a party cannot form a stable voting bloc even though the formation of a bloc would benefit every member because the party faces a collective-action problem: Each member individually benefits more by staying out of the bloc and letting others coordinate their vote, even though they all become better off if they all commit to form a voting bloc. With respect to polarization of preferences, I find that party discipline becomes increasingly difficult to sustain as a stable outcome 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.

Voting blocs are not only a consequence of political parties and their sophisticated partisan strategies. Rather, the coordination 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 chapter I consider an assembly in which any subset of voters can coordinate and coalesce to form a voting bloc. I show that given the configuration into blocs by the rest of the assembly, any arbitrary coalition of agents who form a voting bloc attains a net gain in the sum of expected utilities of its members. I analyze the endogenous formation of voting blocs in the assembly and I seek voting bloc structures -partitions of the assembly into voting blocs- that are stable. I show that there exist Nash stable voting bloc structures. In these structures, no agent has an incentive to leave the bloc she belongs to and join some other bloc. I find that Nash stable voting blocs must be of size less than minimal winning.

To obtain sharper predictions about the configuration of voting blocs, I apply the model to a small assembly and I introduce a new "Split stable" concept that allows for coalitional deviations in which at most one bloc splits. I show that in a stylized assembly with 9 members whose types are symmetrically distributed, all Nash and Split stable voting bloc structures have two voting blocs one at each side of the ideological spectrum and a group of independents including the median in between the two blocs. In the last section of the chapter I compare this result with the predictions derived from empirical data on the voting patterns of the United States Supreme Court from 1995 to 2004.

Using data on the 419 non-unanimous decisions that the Court reached in this period, I provide estimates of the ideal position of each justice in one- and two-dimensional spaces, and I calculate how the formation of voting blocs would have changed the decisions of the Court. For each hypothetical connected voting bloc structure I find the decisions that would have been reversed due to the coordination of votes inside the blocs if this voting bloc structure had formed. I assume that the justices that dissented (voted with the minority) on a decision would have liked a reversal of the decision, and those who in the data voted with the majority and won would have been worse off had the decision been reversed. Aggregating over all the decisions, I calculate the net balance of beneficial minus detrimental reversals for each justice induced by the given voting bloc structure, relative to the original data. Assuming that these individual net balances of reversals are payoffs to the justices, I calculate which connected voting bloc structures are stable if justices form voting blocs to maximize their payoff. The only Nash and Split stable connected partitions involve two voting blocs of size three: Three of the four liberal justices (Stevens, Ginsburg, Souter, Breyer) in a liberal bloc, and the three most conservative justices, namely Rehnquist, Scalia and Thomas in a conservative voting bloc. This empirical exercise shows that justices have strategic incentives to coalesce into voting blocs.

The theory in this chapter draws inspiration from several literary subfields.

In the coalition formation literature, the seminal work of Buchanan and Tullock [12] analyzes the costs and benefits of forming a coalition and praises the virtues of unanimity as internal voting rule. Hart and Kurz [29] study the endogenous formation of economic coalitions under the restriction that the overall partition of the society into coalitions is efficient. Carraro [14] surveys more recent non-cooperative theories of coalition formation, but mostly with economic and not political applications. Traditional models of coalition formation assumed that agents only care about by the coalition they belong to, not by the actions of other agents outside their coalition. The newer *partition function* approach recognizes that agents are affected by the actions of outsiders, and it defines utilities as a function of the whole coalition structure in the society. Bloch [10] and Yi [59] survey the literature on coalitions that generate positive externalities to non-members, such as pollution-control agreements, and coalitions that create negative externalities to non-members, such as custom unions. However, there is no literature on the more general case in which a coalition generates both positive and negative externalities to non-members. The formation of a voting bloc or a political party generates positive externalities to those who agree with the policies endorsed by the party, and negative externalities to agents with an opposed policy preference. My model 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 agent.

In previous formal theories of party formation, Snyder and Ting [57] describe parties as informative labels that help voters to decide how to vote, Levy [38] stresses that parties act as commitment devices to offer a policy platform that no individual candidate could credibly stand for, Morelli [42] 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. Baron [7] and Jackson and Moselle [33] 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. My theory shows that legislators also have an incentive to form parties -voting blocs- in the absence of a distributive dimension, merely to influence the policy outcome over which they have an ideological preference.

From the applied American Politics literature, Cox and McCubbins [15] find that legislators in the majority party in the US Congress use the party as means to control the agenda and the committee assignments, and Aldrich [1] 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. I complement their explanations proving that voting blocs of size less than minimal winning also influence the outcome even if they are not big enough to guarantee a majority, and they generate an ideological policy gain to their members.

Two recent papers in the political economy literature deal with the selection of the voting rule for a single coalition: Barberá and Jackson [6] define *self-selecting* rules as those that would not be beaten by any other rule if the given voting rule is used to choose among rules; Maggi and Morelli [39] study *self-enforcing* rules such that agents would want to undertake collective action under such rule. Another strand of literature studies the formation of governments by coalitions of parties. Four decades after Riker [49] showed the advantages of forming minimal winning coalitions, Diermeier and Merlo [18] show that if agents bargain over ideology and not just the distribution of resources, coalitions may be smaller or larger than minimal winning and in a recent book, Schofield and Sened [54] survey the latest theoretical and empirical findings about the formation of government coalitions in multiparty democracies. Finally, the voting power literature exemplified by the work of Gelman [28] takes a different approach on coalition formation and assumes that agents want to maximize the probability of being pivotal in the decision, instead of maximizing the probability that the outcome is favorable to their interests.

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.

4.2 Motivating Examples

In this section I present three examples to illustrate how the formation of voting blocs affects voting results and policy outcomes. After a simplistic example that illustrates how voting blocs work, I present a more complete example in which two stable voting blocs form in a small assembly, and I present an application to a larger assembly.

Example 3 *Suppose there is an assembly with five agents who have to take a binary choice decision -to approve or reject some action- by simple majority. Suppose that the agents face uncertainty about preferences, in particular, the probability that an agent i favors the action is $\frac{1}{2}$ for each i, and these probabilities are independent across agents. 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 a given 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, such that all three members vote according to the preferences of a majority of members of the bloc -that is, if two members agree, the third votes with them regardless of her own preference. Then the decision reached by the assembly depends exclusively on the preferences of the members of 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 $\frac{3}{4} = \frac{12}{16} > \frac{11}{16}$. Hence, the agents who form a voting bloc increase the probability that the policy outcome coincides with their wishes. The probability for non-members drops to $\frac{8}{16}$.

A bloc of three agents in Example 3 is stable in the minimal sense that no member wants to leave the bloc. If a member leaves, the new situation with a bloc of size two is identical to the original situation with no blocs, because a bloc of two agents is always ineffective: Either both members agree and vote together as they would in the absence of a bloc, or if they disagree, no side holds a majority so each agent is free to vote her true preference, just as if they were not in a bloc.

However, the bloc with three members is not stable if outsiders are free to join in. Indeed, both outsiders want to join. If one or both of them join, the probability that any agent in the assembly obtains her desired outcome becomes $\frac{11}{16}$, which represent a loss for the three original members of the bloc, but a gain to the entrant. The outsiders cannot achieve anything by forming a new bloc of their own because the old bloc of size three is big enough to act as a dictator in an assembly of size five. Some intuitions gained in this example generalize, as I shall show below: Forming a voting bloc always generates a gain in aggregate utility to its members (proposition 32) relative to remaining independent, but if entry to the bloc is open to outsiders, stable blocs cannot be too big relative to the size of the assembly (proposition 38).

In Example 3, the agents are identical random voters, so that only the size of the bloc matter, and not the characteristics of its members. In the rest of the chapter I study 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.

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. Three legislators may sign a pact today to vote together in votes to come in the future, but they do not know today the agenda or the details of the policies they will vote on in the future.[1] 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 random voting model used in Example 3 and in much of the voting power literature[2] is an extreme case of uncertainty, when agents not only do not know exactly how they will feel about future policy proposals, but they cannot even take a guess. In my model, 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. For instance, it is not a foregone conclusion that a Republican legislator in the US Senate will vote in favor of future tax cuts and a Democratic senator against them, but

[1] For instance, the countries of the European Union regularly discuss the notion of a common foreign policy. If some day they sign a treaty establishing a binding common foreign policy, they will sign the treaty with incomplete information about the foreign policy issues that will be salient after the treaty is ratified and comes to effect.

[2] Within this literature, see Felsenthal and Machover [27] for a study of voting blocs, and Hosli [31] and [32] for an exception in which she calculates the voting power of different states in the European Union council taking into account preferences by noting that some coalitions are more likely to emerge than others.

it is ex ante more probable that the Republican, rather than the Democrat, will favor the tax cuts.

The ex ante differences in the preferences of the agents are key determinants of the strategic incentives to form voting blocs. Let us see how a polarized small assembly splits into two different voting blocs, none of which is minimal winning.

Example 4 *Suppose there is an assembly with seven agents who have to take 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 t_i. Suppose $t_1 = t_2 = t_3 = \frac{1}{4}$ and $t_4 = t_5 = t_6 = t_7 = \frac{3}{4}$. Table 4.1 shows the probability that the policy proposal gathers at least four votes and passes unconditional (column one) and conditional on agent* 7 *favoring the policy proposal (column two), and the probabilities that the outcome coincides with the preferences of agents* 7 *(column three),* 4 *(column four) and* 1 *(column five), given that the following voting blocs form: no blocs (row one); agents* 5, 6, 7 *form a bloc (row two); agents* 1, 2, 3 *form a bloc (row three); both blocs form (row four); agents* 1, 2, 3 *form a bloc and* 4, 5, 6, 7 *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.*

Table 4.1: Probabilities that the proposal passes and the agents like the outcome

Bloc	*Pass*	*Pass| 7 favors*	*7 satisfied*	*4 satisfied*	*1 satisfied*
None	59.4%	68.8%	68.8%	68.8%	57.3%
{5,6,7}	75.7%	83.9%	75.3%	76.3%	41.9%
{1,2,3}	42.3%	51.2%	59.5%	59.5%	63.4%
{1,2,3},{5,6,7}	68.4%	74.7%	68.6%	85.4%	45.4%
{1,2,3},{4,5,6,7}	77.1%	86.6%	77.8%	77.8%	39.4%

The numbers on the table come from simple binomial calculations. Glancing at the table it is evident that the formation of different voting blocs has a significant effect on the outcome. Note that regardless of whether the three low types form a bloc or not, three high types are better off forming a voting bloc if they take the actions of the other members as given, and similarly, given that a bloc with three high types form, or given that it does not form, the three low types are better off forming their bloc. The outcome with two blocs of size three is Nash stable -no member of a bloc wants to leave, no other agent wants to enter.

Note that agent 4 in Example 4, identical in all respects to agents $5, 6, 7$ does not want to join the bloc of high types. Rather, with two opposing blocs the remaining independent agent is better off, since the two blocs are likely to counterbalance each other and the outcome is then often left for the independent to decide. With only the purely ideological motivation of caring for the policy outcome and no rents from office to distribute among the members of the winning coalition, agent 4 has no incentive to join the bloc of high types, and blocs will not be of minimal winning size.

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. For ease of calculation and exposition, the agents in the following example have very specific preference profiles; this is only for illustration, and the body of the chapter generalizes the intuitions provided in the example.

Example 5 *Consider an Upper House with 100 members in a bicameral system such that a bill approved in the Lower House requires 51 favorable votes in the Upper House to become Law, otherwise a status quo remains in place. Suppose that the Lower House is under liberal control and always passes liberal bills, while legislators in the Upper House come in three types:* 20 *conservatives,* 30 *moderates and* 50 *liberals. Conservatives oppose every bill, liberals favor every bill, and each moderate favors exactly one third of the bills, in such a way that exactly one third of the moderates favor any given bill. Then, in the Upper House every bill has* 60 *supporters (*10 *moderates and* 50 *liberals) and* 40 *opponents (*20 *conservatives and* 20 *moderates) and in the absence of voting blocs the advocates of the bill always win.*

Suppose all 30 *moderates and* 10 *conservatives in the Upper House form a voting bloc which they call the Coalition, and they commit to always vote together by first meeting in a Coalition Caucus and reaching a common position by simple majority in the caucus. For each bill that comes to the*

Upper House, the Coalition Caucus gives the same outcome: $10 - 30$ *against the bill. Then, if all the members follow the dictates of the voting bloc they have just formed, in the division of the Upper House none of them votes for the bill, each bill then receives only* 50 *votes in the floor of the Upper House, all coming from liberal legislators, and the bill is defeated.*

The formation of a voting bloc by a minority in Example 5 crucially affects voting results, policy outcomes, the utilities of the legislators involved and, ultimately, the utility of their constituencies. Conservative legislators now always achieve their desired outcome (defeating the bill). Moderate legislators achieve their desired policy in two out of three cases (those in which they oppose the bill), which is twice as often as without a voting bloc. But the example shows only the potential gains of forming a bloc, not the difficulties in making it stable to safeguard these gains. The minoritarian Coalition dominates the Upper House in Example 5 because it manages to forge a voting bloc that quashes internal dissent and shows no fissures in voting patterns.

The Coalition is not stable. Every moderate has an individual incentive to leave. Suppose one moderate defects and becomes an independent. If the independent opposes the bill, the defection has no effect; the bill gathers only 50 votes and fails. But in the event that the independent favors the bill, the bill passes $51-49$. The independent is now pleased with the outcome with certainty, thus she benefit from her defection to the detriment of those legislators who remain in the Coalition.

The probabilities over events in this example are contrived to make calculations trivial, but two important intuitions generalize.

First, note that while every moderate has an incentive to abandon the Coalition, the conservatives gain nothing by leaving. In proposition 26 below I show that given a voting bloc that leans towards one side of the political spectrum, it is only the most moderate members of the bloc who threaten the stability of the bloc; if the moderate members benefit from participating in the bloc, it follows that the extreme members also benefit from participating. In other words, it is only the liberal wing of a conservative party, and the conservative wing of a liberal party who determine whether the party can form a stable voting bloc. Intuitively, if the most left-leaning legislator in the US House of Representatives introduces a bill to her liking, the hope that it passes must lie in gaining the favorable vote of Democrats, as it is implausible that a progressive liberal bill could pass against the opposition of Democrats by gathering enough Republican votes. If a majority of Democrats opposes the bill, the bill is doomed anyway, so the legislator has nothing to gain in terms of policy outcomes by defecting from the Democratic Party.

Second, suppose that the Coalition Caucus changes its rules and adopts the following supermajority internal voting rule: Members have freedom to vote in the Upper House according to their own wishes unless three quarters of the members of the Coalition share the same view, in which case the whole Coalition must vote together. In a bloc of either 39 or 40 members this rule requires that at least 30 members share the same opinion before the minority is forced to vote with the majority. If all moderates stay in the Coalition they always achieve the threshold and the bloc functions equally as if it was using simple majority: Thirty members oppose the bill, so the whole Coalition votes against it and the bill fails with just 50 liberal votes. Now consider the incentives of a moderate given the new rule. As a member of the bloc, the legislator attains the outcome she wants whenever she opposes the bill, which occurs with probability two thirds. Suppose she leaves the bloc and she opposes the bill. Then there are only 29 legislators left opposing the bill inside the bloc, not enough according to the new rule to force the minority of dissenters to reverse their vote. Hence 10 moderates vote for the bill, and the bill passes $60-40$. The deviant is now worse off as an independent, because her vote is necessary for the Coalition to act together as a voting bloc, and as a result the Coalition with the new supermajority rule becomes stable.

This result generalizes, as shown in proposition 29: Under weak conditions on the size of the parties that compose the assembly, and for any size of the assembly, there exist type profiles such that a party cannot form a stable voting bloc if it chooses simple majority as its internal voting rule, but it can form a stable voting bloc with some supermajority internal voting rule.

The first preliminary insight into the gains reaped by voting blocs is the following: Whenever the 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.

The second basic insight is that generating a gain is not sufficient for the bloc to be stable -or to form in the first place. Rather, it must be that every agent has a strategic incentive to participate in the bloc. The rest of the chapter explores the individual incentives to participate in blocs, and the resulting stability properties of different voting bloc formations as a function of the preferences of the members of the assembly and of the voting rules used by the voting blocs.

4.3 The Model

Let $\mathcal{N}$ =$\{1, 2, ..., N\}$ be an assembly of voters, where $N \geq 7$ is finite and odd. This assembly must take 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 in favor or against the proposal and zero otherwise, hence there is no intensity of preferences. Let $s_i = 1$ if agent i prefers the proposal to pass, and zero otherwise; let $s = (s_1, ..., s_N)$ be a preference profile for the whole set of voters, and let $s_{-i} = (s_1, ..., s_{i-1}, s_{i+1}, ..., s_N)$ 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.

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 $s \in S = \{0, 1\}^N$, the probability that s will occur. Let $\Omega : \{0, 1\}^N \longrightarrow [0, 1]$ be the probability distribution over profiles and assume Ω is common knowledge. Let t_i be the type of agent i, which is the probability that i favors the policy proposal, so $t_i = P[s_i = 1]$. If these probabilities are not correlated across agents, then I say that types are independent.

Definition 10 *Types are independent if* $\Omega(s) = \prod_{i \in \mathcal{N}} [t_i s_i + (1-t_i)(1-s_i)]$ *for all* $s \in S$.

If types are independent the probability that i favors the policy proposal is t_i for any given realization of preferences by the rest of agents, that is, $P[s_i = 1 | s_{-i}] = t_i$ for all $i \in \mathcal{N}$ and all $s_{-i} \in S_{-i} = \{0,1\}^{N-1}$.

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 $\mathcal{N} = L \cup R \cup M$. 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. This setting applies to a legislature such as the US House of Representatives or the US Senate in which legislators affiliate to one of the two dominant political parties, or remain independent. Each of the two parties $C = L, R$ can coordinate the voting behavior of its members by forming a voting bloc $V_C = (C, r_C)$ with an internal voting rule r_C that maps the preferences of its members into votes cast by the bloc in the division of the assembly. Then it becomes a strong party that exhibits party discipline in voting. I assume that joining a voting bloc is voluntary, so the party as a whole forms a voting bloc only if every member wants to participate in it, otherwise only a coalition of agents representing the subset of party members who want to participate form a voting bloc, and the rest of party members do not coordinate their votes, effectively becoming independents. Independent agents who are not originally affiliated to a party do not coordinate their votes.

The timing of events is as follows.

Given an arbitrary pair of internal voting rules r_L and r_R, each member in party L simultaneously chooses whether to join the voting bloc with rule r_L or remain outside the bloc to act independently, and similarly each member in R chooses whether to join the voting bloc with rule r_R or not. Two voting blocs thus form, each bloc containing the members of a party who choose to join it. Then a preference profile s is realized, each agent i learns her own preference s_i, and the two voting blocs hold their internal meetings. Finally the whole assembly meets and agents vote according to the outcome of their voting bloc's internal meeting if they have joined any, or according to their own wishes otherwise.

Given a coalition C of size N_C that forms a voting bloc, in their internal meeting the members of C determine their coordinated voting behavior according to their own internal rule r_C, where $r_C = r_L$ if $C \subseteq L$ and $r_C = r_R$ if $C \subseteq R$ and I assume that the voting bloc has commitment mechanisms such that the outcome of this internal meeting is binding. In particular:

1. If $\sum_{i \in C} s_i \geq r_C N_C$, then $\sum_{i \in C} v_i = N_C$. If the fraction of C members who prefer the policy proposal is at least r_C, then the whole bloc votes for the proposal in the division of the assembly.

2. If $\sum_{i \in C} s_i \leq (1 - r_C) N_C$, then $\sum_{i \in C} v_i = 0$. If the fraction of C members who are against the policy proposal is at least r_C, then the whole bloc votes against the proposal in the division of the assembly.

3. If $(1 - r_C) N_C < \sum_{i \in C} s_i < r_C N_C$, then $\sum_{i \in C} v_i = \sum_{i \in C} s_i$. If neither side gains a sufficient majority within the voting bloc, each member votes according to her own preference in the division of the assembly.

I assume that r_L and r_R are such that the thresholds $r_L N_L$ and $r_R N_R$ are integers weakly larger than $\frac{N_L+1}{2}$ and $\frac{N_R+1}{2}$ respectively. With an $r_C - majority$ internal rule, the integer $r_C N_C$ is the number of votes the majority in the voting bloc (C, r_C) must gather in order to roll the internal minority and act as a unitary player in the division of the assembly, casting all N_C votes in favor of the policy advocated by the majority of the bloc. A rule $r_C = \frac{N_C+1}{2N_C}$ is simple majority, and $r_C = 1$ is unanimity, which is identical to not coordinating any votes -members only vote together if they all share the same preference.

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, sincere voting is weakly dominant; voting against her preference can only make an agent worse off. Therefore, it is safe to assume that members of a bloc reveal their preference truthfully and then it is without ambiguity that I use the same notation for the true preference s_i and the vote of agent i inside the bloc.[3] In a more straightforward interpretation that bypasses internal voting, 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.

Since non-dominance alone results in sincere voting, the only remaining strategic consideration in the model is about membership in a voting bloc. Participation in a voting bloc is voluntary, and members of a party choose to join their party's voting bloc according to their own individual incentives. Agents seek to maximize the ex ante (before preferences are revealed) probability that the policy outcome in the assembly coincides with their policy preference. They only participate in a voting bloc if belonging to the bloc increases this ex ante probability.

I seek to explain under what conditions a party can behave as a cohesive unit, forming a stable voting bloc in which every member voluntarily participates. While the equilibrium properties of the entry game I have described are interesting, I focus on the narrower question of the stability of the party. Rather than searching for equilibria after the original parties break up and subsets of these parties form voting blocs, I find under what conditions a party can form a voting bloc imposing voting discipline on its members and every member accepts the party discipline so that the voting bloc is stable. The stability concept that I use is merely a voluntary participation condition. Whoever belongs to a bloc must be weakly better off as a member of the bloc than deviating to become an independent. If a single party member does not want to join the bloc given that every other member does, the party cannot form a stable voting bloc with voluntary participation.

[3] To be formally precise, I would need to define a new variable $\widehat{s}_i$ to denote the preference expressed by i in the internal meeting of coalition C, and let $\sum_{i \in C} \widehat{s}_i$ determine the outcome of the internal meeting, but since sincere voting is weakly dominant, $\widehat{s}_i = s_i$ for all $i \in C$, all $C \subseteq \mathcal{N}$ and all $s \in S$.

Definition 11 *A voting bloc $V_C = (C, r_C)$ is Individual-Exit stable if every member $i \in C$ weakly prefers to join the bloc than to become an independent and let the bloc $(C \backslash i, r_C)$ form instead.*

Members of a voting bloc must be weakly better off ex ante, at the time they commit to participate in the bloc, before they learn their own preferences. Once voting blocs form, I assume that there are binding mechanisms so that ex post the losing minority within a bloc cannot renege from the commitment to vote with the bloc's majority; that is, the outcome of the internal meeting of the voting bloc is enforced.

This is a partial-equilibrium definition: For $C = L, R$, a voting bloc $V_C = (C, r_C)$ is Individual-Exit stable if it is a best response in the entry game for every agent in C to join the voting bloc given that every other agent in C joins the bloc, and taking as given the outcome of the formation process in the other party. Each party may then be Individual-Exit stable conditional on the formation or not of a voting bloc in the other party. I study the stability of the assembly as a whole at the end of the section, providing a more formal extension of Definition 11 to encompass the incentives to deviate by all agents in multiple voting blocs. First, I focus on the formation of a bloc by a given party as a best response to the strategies of the other party.

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. To make this informal statement more precise, I introduce some notation.

Given the probability distribution Ω over preference profiles, for any $C \subseteq \mathcal{N}$, let $g^C(x)$ be the probability that $\sum\limits_{i \in C} s_i = x$. For any i, h in C, let $g^C_{-i}(x)$ be the probability that $\sum\limits_{k \in C \backslash i} s_k = x$ and let $g^C_{-ih}(x)$ be the probability that $\sum\limits_{k \in C \backslash \{i,h\}} s_k = x$.

Definition 12 *A set of agents C of odd size N_C leans right if for any non-negative k*

$$g^C(\frac{N_C - 1}{2} - k) \leq g^C(\frac{N_C + 1}{2} + k).$$

C leans left if the inequality signs are reversed and is symmetric if the condition holds with equality.

For a set of even size, the relevant inequalities are:

$$g^C(\frac{N_C}{2} - k) \leq g^C(\frac{N_C}{2} + k) \text{ for any positive } k.$$

Definition 12 is best interpreted as follows: A coalition C leans right if for any size of the internal majority and minority within the coalition, it is at least as likely that the majority favors the policy proposal than that the majority rejects the proposal. Similarly, if for any majority-minority split of preferences it is more likely that the majority rejects the alternative, then the coalition leans left.

Definition 12 extends the definition of "leaning towards a" in chapter 3 to make it applicable to any subset of agents, and it provides a natural one-dimensional interpretation in which we imagine that the proposal is always conservative (right) so that left-leaning agents often reject the proposal and right-leaning agents tend to accept it. Formally, coalition C in chapter 3 "leans towards a" if and only if the subset $C \backslash \{i\}$ leans right for any $i \in C$.

Assuming that one party leans left, a second party leans right, and the set of independent agents is symmetric, the following results show the necessary and sufficient condition on the types of the members of a party for this party to be able to form a stable voting bloc, given that the opposing party forms (or does not form) its own bloc.

Lemma 25 *Let $\mathcal{N} = L \cup M \cup R$. Suppose types are independent, L leans left and forms a voting bloc (L, r_L), M is symmetric and for any $i \in R$, R_{-i} leans right. Let $l \in R$ be such that $t_l \leq t_i$ for all $i \in R$. Suppose R forms a voting bloc with an internal rule r_R. If agent l prefers to participate in the voting bloc (R, r_R) than to become an independent, then every i in R prefers to participate.*

Lemma 25 provides an important insight: The stability of the voting bloc (R, r_R) depends only on the agent with the lowest type in R. The intuition is that if the most leftist member in party R has an incentive to participate in a right-leaning voting bloc, then every other party member has an even greater incentive to belong to the bloc. The left-most agent is the least likely to benefit from the actions of the bloc and the most likely to be rolled to vote against her wishes, hence if she doesn't want to deviate, no one else will.

Lemma 25 and other results below assume that M is symmetric and L leans left and forms a voting bloc (L, r_L). This assumption can be weakened. First, since the result holds if r_L is unanimity, it implicitly holds as well if no voting bloc forms in L -since forming a voting bloc with unanimity is identical to not forming a bloc. More generally, it suffices to assume that the distribution of votes cast in the division of the assembly by the set of agents $L \cup M$ (those not in R) is such that given any size of the majority vote in $L \cup M$, with probability at least $\frac{1}{2}$ this majority is against the proposal. Formally, it suffices that:

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

This condition is similar to $L \cup M$ leaning left, but it applies to the probability distribution of actual votes cast in the division of the assembly after accounting for the coordination of votes inside L, and not to the probability distribution over true preferences.

Lemma 25 shows that inside each party only one extreme agent matters to determine whether the party can form a stable voting bloc. In particular, I next show that a party leaning right can form a stable voting bloc if and only if its left-most agent is not too far to the left, or, in other words, if the lowest type in the party is high enough.

Proposition 26 *Let $\mathcal{N} = L \cup M \cup R$. Suppose that types are independent, L leans left and forms a voting bloc (L, r_L), M is symmetric and for any $i \in R$, R_{-i} leans right. Let $l \in R$ be such that $t_l \leq t_i$ for all $i \in R$. Suppose R forms a voting bloc with an internal rule r_R. The voting bloc (R, r_R) is Individual-Exit stable if and only if t_l is higher than a cutoff function $t^{InR}(r_R, r_{L,} t_{-l})$.*

If the type of agent l is high enough, she wants to participate in the (R, r_R) voting bloc; if she wants to participate, every other member of R wants to participate and the bloc is Individual-Exit stable. Intuitively, if R forms a voting bloc the main consequence is that with a high probability most members of R favor the policy proposal, those who do not are rolled and compelled to vote in favor of it, and the policy proposal passes with a higher probability. An agent l only wants to join such a bloc that essentially turns around naysayers to make them support the policy proposal if l likes the proposal with a high enough probability. The cutoff is a function of the sizes of the blocs, the internal rules they use, and the types of all the other agents in the society.

A symmetric result applies to the left party. Given that (R, r_R) forms, the voting bloc (L, r_L) is stable only if the highest type t_h among the members of L is below a cutoff function $t^{InL}(r_L, r_R, t_{-h})$. Taking both results together, a corollary follows:

Corollary 27 *Let $\mathcal{N} = L \cup M \cup R$. Suppose that types are independent, L_{-i} leans left for all $i \in L$, M is symmetric and R_{-j} leans right for any $j \in R$. Let $h \in L$ be such that $t_h \geq t_i$ for all $i \in L$ and let $l \in R$ be such that $t_l \leq t_j$ for all $j \in R$. Then it is a Nash equilibrium of the entry game for every agent in L and R to respectively join (L, r_L) and (R, r_R) if and only if*

$$t_h \leq t^{InL}(r_R, r_{L,} t_{-h}) \text{ and } t_l \geq t^{InR}(r_L, r_{R,} t_{-l}).$$

The two parties can each form a stable voting bloc if the highest type in the left bloc is not too high, and the lowest type in the right bloc is not too low. Note that the types of the members of each bloc may overlap, i.e., the right-most member of the Left bloc may be to the right of the left-most member of the Right bloc, but not too far to the right, and similarly agents too far to the left will not belong to the Right bloc.

The exact threshold $t^{InR}(r_R, r_L, t_{-l})$ above which an agent with type t_l wants to join the voting bloc (R, r_R) depends on the size and voting rule of the bloc (L, r_L), the number of independents and the type profile of all agents other than l in the assembly, all of which are variables exogenous to R. But it also depends on party R, both on its size and the voting rule it uses to aggregate preferences within its own bloc.

Simple majority, $r_C = \frac{N_C+1}{2}$, is the internal rule that maximizes the sum of utilities of the members of a voting bloc $V_C = (C, r_C)$. I show this in detail in the more general proposition 32 below, but the intuition is as follows: A voting bloc only has an effect in utilities if the coordination of the voting behavior of its members alters the policy outcome in the division of the assembly. A voting bloc subtracts votes from the position supported by its internal minority, adding them to the internal majority position. Hence, if the bloc alters the outcome in the assembly, it changes it from the outcome preferred by a minority of the members of the bloc to the one preferred by a majority of members of the bloc. Since there is no intensity of preferences, it follows that the sum of utilities in the bloc increases. Simple majority always rolls the minority votes, so it maximizes the probability that the bloc alters the outcome in the division of the assembly and gains a surplus, so it maximizes the sum of utilities of the bloc.

Notwithstanding the advantages of simple majority, for some parameters a bloc with simple majority is not stable: Agents face a temptation to leave and "free ride" from the coordination of votes by the bloc. Other supermajority r_C internal rules reduce the surplus gained by the bloc, but in some instances make the bloc stable. I explore this possibility in the following two results.

Proposition 28 *Let* $\mathcal{N} = L \cup M \cup R$. *Suppose types are independent and* R *is composed of* N_R *homogeneous agents with a common type* t_R. *Then the voting bloc* (R, r_R) *with* $r_R = \frac{N_R - 1}{N_R}$ *is Individual-Exit stable.*

The voting bloc thus formed is stable regardless of the formation process of other voting blocs or the types of other agents in the assembly. Indeed, irrespective of the other agents, forming a bloc generates a surplus for its members, as discussed briefly above and proved below for a more general case in proposition 32. Identical members of a homogeneous bloc all benefit from the surplus. Under a supermajority rule $\frac{N_R - 1}{N_R}$, if the bloc loses a single member, it effectively disbands, since with the reduced membership it would only reach the internal threshold for a sufficient majority if all agents agree, in which case the bloc never affects the outcome and generates no surplus. For example, imagine a bloc with 7 members and a 6/7 rule, so that only minorities of one are rolled. If an agent deviates and leaves the bloc, the new bloc with 6 members and a 6/7 rule is irrelevant: A majority of 5 to 1 does not represent a 6/7 majority, so the bloc never rolls its minorities. Thus, a stringent supermajority rule that makes every agent essential to roll a minority deters exit -at least in a homogeneous bloc. The loss in surplus is significant with such a stringent rule, since the bloc forsakes the opportunity to roll bigger minorities granted by simple majority. However, as shown in the next proposition, there are some parameters for which a bloc with simple majority is not stable, and if a party wants to form a stable bloc, it would need a more stringent internal voting rule.

Proposition 29 *Let $\mathcal{N} = L \cup M \cup R$. Suppose that types are independent with $t_i \in (0,1)$ for all $i \in \mathcal{N} \backslash R$. 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 exists a vector of type profiles for the members of R such that R leans right and (R, r_R) is not Individual-Exit stable if r_R is simple majority, but it is Individual-Exit stable for some supermajority internal rule r'_R.*

Parties that cannot form a voting bloc with simple majority (because their members would leave), can form a stable voting bloc that only coordinates the votes of its members when the internal majority in the party is more substantial than a mere majority of one. Figure 4-1 illustrates this result. To be able to plot $t^{InR}(r_R, r_L, t_{-l})$ as a function of a single parameter, I let $N_L = 11$, $N_R = 9$, $N_M = 31$, $t_i = 0.3$ for all $i \in L$,[4] $r_L = \frac{6}{11}$ (simple majority), $t_m = 0.5$ for all $m \in M$ and $t_h = t_R$ for all $h \in R \backslash l$. Given these values, I plot $t^{InR}(r_R, r_L, t_{-l})$ as a function of t_R for $r_R = \frac{5}{9}$ (simple majority), 2/3 and 8/9. It is clearly observed that the more stringent the internal voting rule of R, the lower the type of l can be such that l wants to participate in the voting bloc (R, r_R).

While I do not study in this work the endogenous selection of voting rules for a party, it follows that 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 the voting bloc be Individual-Exit stable is the lowest possible rule such that the bloc is stable. For some parameters, this rule is

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

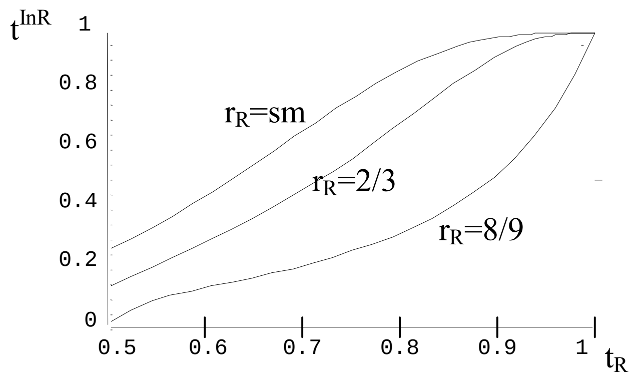

Figure 4-1: Individual-Exit stability of (R, r_R) for $r_R \in \{1/2, 2/3, 8/9\}$.

a supermajority.

This result contrasts with the findings of Maggi and Morelli [39] who study a single coalition that votes on whether or not to take a collective action. They find that the optimal self-enforcing 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. A key difference between this model and theirs is that they restrict attention to homogeneous agents (or in their terminology, "symmetric" agents) who all share the same type. A second important difference is that in their model the collective action of the coalition does not generate an externality to non-members. I show that once we take into account that agents are heterogeneous and that the actions of a coalition generate externalities to non-members, 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 join a voting bloc 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 [53] and Aldrich and Rohde [2], who look at party discipline in the US Congress and conclude that backbenchers 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 [15], 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 29 shows that this finding is not an idiosyncrasy of the Democratic and Republican parties in the US Congress, but rather, a general principle is 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.

Next I study how the possibility that a party forms a stable voting bloc depends on the extremism of the types of its members. I show that a sufficiently extreme party cannot form a stable voting bloc unless it uses the very restrictive almost-unanimity internal voting rule considered in proposition 28.

Proposition 30 *Let* $\mathcal{N} = L \cup M \cup R$. *Suppose types are independent,* M *is symmetric,* R *leans right and forms a voting bloc* (R, r_R) *and* $N_L < N_R + N_M$. *Let* $(x_1, ..., x_{N_L})$ *be an arbitrary vector such that* $x_i \in [0,1]$ *for all* $i = 1, ..., N_L$. *Suppose the types* $(t_1, ..., t_{N_L})$ *of the members of* L *are of the form* $t_i = \alpha x_i$. *If* α *is low enough,* (L, r_L) *with* $r_L \leq \frac{N_L - 2}{N_L}$ *is not Individual-Exit stable.*

A corollary to proposition 30 is that no extreme party of size more than three but less than minimal winning can form a voting bloc with simple majority, even if its members all share a common type. The intuition for this negative result on the stability of 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 being rolled when they happen to favor the policy proposal. Only if $r_C N_C = N_C - 1$ the result in proposition 28 applies and a party of extremist is stable because if one of them steps out of the bloc, the bloc dissolves and there is no possibility to enjoy the benefits of the formation of the bloc while remaining out of it.

With simple majority as internal voting rule, the maximum size up to which a minority voting bloc is stable is inversely related to the extremism of its members. I show this in a numerical example, which tracks the maximum size of parties capable of forming stable voting blocs as a function of the polarization of both a symmetric and an asymmetric assembly, split into two homogeneous parties one at each side of the political spectrum and a number of moderate independents in between.

Example 6 *Let* $\mathcal{N} = L \cup M \cup R$*. 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* (L, r_L) *and* (R, r_R) *are Individual-Exit stable 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 example illustrates the plausible intuition that extremists are only able to coordinate in small numbers, while moderate agents can form larger voting blocs.

Sometimes a party cannot form a stable voting bloc 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 have an incentive to leave the party and let others coordinate their votes. To address this collective action problem, suppose that a party does not allow a single member to step out of the bloc and free ride on the advantages provided by the voting bloc, but rather, the party only forms a voting bloc if all of its members participate. In other words, the party renounces to enforce any voting discipline if a single member refuses to accept it.

If a party is able to commit to an "all or none" outcome in which either the whole party forms a voting bloc or every party member votes independently, individual incentives to participate change: Now an agent cannot leave the party and expect to reap the benefits from the rolling of minority votes inside the bloc while facing no risk of ever being forced to vote against her own preference. In the new calculation each agent weighs the gain brought by the bloc, and not the marginal advantage of being in or out of a bloc that forms. It follows that under some circumstances, agents who would prefer not to participate in the bloc now choose to join only because their participation becomes essential to the very existence of party discipline. By committing to form a bloc only by unanimous agreement, a party can sometimes overcome the collective action problem it faced under individual participation.

Given that the opposite party forms a voting bloc, suppose party $C \in \{L, R\}$ plays the following game $\mathcal{G}_C$: Every $i \in C$ simultaneously chooses whether or not to sign a conditional participation contract, by which i joins the voting bloc if and only if every other member in C joins the bloc too. If every $i \in C$ signs the contract, then the party forms the voting bloc (C, r_C), otherwise member of C do not coordinate their votes.

The players in the closed membership game $\mathcal{G}_C$ are the N_C members of party C. The set of pure strategies of each player is binary: Sign or not sign. Payoffs for each agent are given by the probabilities over policy outcomes determined by Ω and by the voting blocs resulting from the game. The internal voting rule for the bloc is in this description exogenous, but it could be incorporated into the strategy of the players, making the party form a bloc if and only if all the members agree on a common rule, and those who do not want to participate can merely propose unanimity, assuring that no bloc with a rule different than unanimity can form. Let $\mathcal{G}$ describe a larger game in which both parties choose simultaneously whether or not to form a voting bloc by signing conditional contracts.

Results similar to lemma 25 and proposition 26 apply to the closed membership game $\mathcal{G}_C$ just described (these results are available from the author). If the type of the member with the lowest type in party R is high enough then this member benefits from the formation of the bloc (R, r_R) and it is a weakly dominant strategy for her to commit to participate in the voting bloc. If so, it is a weakly dominant strategy for every member to commit to participate.

Party C has something to gain by playing the game $\mathcal{G}_C$ to form a voting bloc instead of trying to form a bloc with whoever wants to join. For some parameters, by threatening not to form a bloc if a single party member fails to join, in the equilibrium of the game $\mathcal{G}_C$ party C forms the voting bloc (C, r_C) even though this bloc is not Individual-Exit stable if the bloc does not dissolve after an individual deviation. Proposition 31 states this result formally.

Proposition 31 *Let $\mathcal{N} = L \cup M \cup R$. Suppose that types are independent, M is symmetric, R leans right and forms a voting bloc (R, r_R) and $N_L < N_M + N_R$. Then there exists a vector of types $(t_1, ..., t_{N_L})$ for the members of L for which the voting bloc (L, r_L) with $r_L \leq \frac{N_L - 2}{N_L}$ is not Individual-Exit stable, but in the game $\mathcal{G}_L$ it is a Nash equilibrium in weakly undominated strategies for every $i \in L$ to commit to participate in (L, r_L).*

In short, if the dissolution of the voting bloc follows the departure of a single party member, then such departure -which would occur if the bloc did not react to the deviation and continued functioning with a shrunk membership- is forestalled. This result extends to the game $\mathcal{G}$ in which both parties use conditional contracts to determine the formation of their respective voting blocs: There exist vector of types for which, regardless of whether R forms or not a voting bloc, party L cannot form an Individual-Exit stable voting bloc with $r_L \leq \frac{N_L - 2}{N_L}$, but nevertheless in the Nash equilibrium of the game $\mathcal{G}$ every member of L commits to participate in the voting bloc (L, r_L). Conditional contracts to form a voting bloc only with unanimous participation allow parties to solve the collective-action action problem that sometimes arises when parties try form a voting bloc to coordinate the votes of their members.

I have studied the incentives of each of two parties to form a voting bloc. Proposition 26 shows the necessary and sufficient condition for a voting bloc with a majority rule to be Individual-Exit stable. Propositions 29 and 31 propose two solutions that can help a party form a voting bloc when a bloc with simple majority is not stable: Either to use a supermajority, or to commit not to form the voting bloc unless every party member participates, if such a commitment is possible.

In the following subsection, I generalize the model by weakening several assumptions and allowing new voting blocs to form.

4.3.1 Generalization: Endogenous Voting Blocs

The model so far applies to an established assembly that uses simple majority as its voting rule and has two well-defined parties. The results have shown under what conditions these parties can form stable voting blocs that coordinate the votes of all their members.

Now imagine instead an assembly $\mathcal{N}$ where all agents are free to coalesce with whomever they wish, with no preassigned cleavages or factions to restrict their coordination with any other member of the assembly. The assembly uses a majority voting rule $r_{\mathcal{N}}$ which may differ from simple majority, such that the (still exogenous) policy proposal passes if it gathers at least $r_{\mathcal{N}}N$ votes and a status quo stays in place otherwise.

The probability distribution over preference profiles is as before Ω, but types need not be independent. Rather, the only restriction that I impose for some results below is that Ω has full support, that is, $\Omega(s) > 0$ for all $s \in S = \{0, 1\}^N$.

Agents form voting blocs facing uncertainty over preferences, then they privately learn their own preference, they vote internally in the voting bloc they belong to, and then they vote in the assembly according to the outcome of their bloc or according to their own wish if they are not members of any bloc.

I am interested in the problem of finding a configuration of the assembly into voting blocs that is stable. Let C_0 denote the subset of agents who remain independent and do not coordinate their votes with any other agent. I treat this subset of agents as if they formed a voting bloc with unanimity as its internal voting rule, so that they only vote together if they all agree. Then, I refer to the configuration of agents into voting blocs in the assembly as the voting bloc structure of the assembly.

Definition 13 *A voting bloc structure (π, r) is a pair composed of a partition of the assembly $\pi = \{C_j\}_{j=0}^{J}$ and a corresponding set of voting rules $\{r_j\}_{j=0}^{J}$ such that $r_0 = 1$ and for $j \in \{1, ..., J\}$, $V_j = (C_j, r_j)$ is a voting bloc with internal rule r_j.*

Note that the voting bloc structure specifies both the membership of each voting bloc, and the rule that each bloc uses to aggregate its internal preferences. I assume that for any voting bloc $V_C = (C, r_C)$, the rule r_C is such that $r_C N_C$ is an integer weakly larger than $\frac{N_C+1}{2}$, where N_C is the size of coalition C. When I consider deviations from the voting bloc (C, r_C), I assume that r_C does not change following the defection of some members or the entry of a new member; as a result, in the new voting bloc (C', r_C) with size N'_C that follows the deviation it is possible that $r_C N'_C$ is no longer an integer.

The voting bloc structure (π, r), together with the preference profile s determines the vote of agent i in the division of the assembly, which I denote $v_i(\pi, r, s)$. Let $u_i(\pi, r)$ be the ex ante expected utility for agent i with the voting bloc structure (π, r).

An important result is that any coalition of agents attains a non-negative net change in aggregate utilities if they form a voting bloc instead of remaining independent, regardless of the configuration of the rest of the assembly.

Proposition 32 *Given a voting bloc structure (π, r), suppose a subset $C'_{J+1} \subset C_0$ deviates and forms a new voting bloc (C'_{J+1}, r'_{J+1}). Denote the resulting voting bloc structure in which no further deviations from (π, r) take place by (π', r'). Then $\sum\limits_{i \in C'_{J+1}} u_i(\pi', r') \geq \sum\limits_{i \in C'_{J+1}} u_i(\pi, r)$. A simple majority internal voting rule r'_{J+1} maximizes the sum of utilities for the members of the new voting bloc $\{C'_{J+1}, r'_{J+1}\}$.*

As discussed in the previous subsection, a voting bloc only has an effect if it reverses the policy outcome. If it does, it favors its internal majority at the expense of its internal minority, generating a net gain. With simple majority the bloc always rolls its internal minority, maximizing the probability that the bloc alters the outcome in the division of the assembly and generates the mentioned net gain. Consequently, simple majority is the internal rule that maximizes the sum of utilities of the members of the bloc, just as it is the rule for the assembly that maximizes the utilitarian social welfare if all agents vote independently, as shown by Curtis [16].

Not only the members of the new bloc benefit. Agents whose preference coincides with the majority of the new bloc with high enough probability also benefit from the formation of the bloc. In the aggregate, the impact of a new voting bloc on the utilities of non-members depends on the voting rule $r_{\mathcal{N}}$ of the assembly.

I say that Ω has full support if every profile of preferences s occurs with strictly positive probability.

Proposition 33 *Suppose Ω has full support. Let (π, r) be any voting bloc structure. Let (π', r') be a new voting bloc structure in which a subset C of size at least 3 contained in the original set of singletons C_0 forms a voting bloc with a rule $r_C < 1$, and the rest of the structure remains unchanged. If $r_{\mathcal{N}} = 1$, then $u_i(\pi', r') > u_i(\pi, r)$ for any i in the new set of singletons C_0' and $\sum_{i \in \mathcal{N}} u_i(\pi', r') > \sum_{i \in \mathcal{N}} u_i(\pi, r)$.*

If the assembly uses a unanimity rule, the members of any voting bloc that forms with an internal rule short of unanimity relinquish their veto power. Agents who retain their veto powers by staying out of any bloc then unambiguously benefit. If the assembly did not use unanimity to begin with, the picture is murkier: Some agents will typically benefit by the formation of a voting bloc, others will suffer. Social welfare may increase or decrease depending on whether the new structure makes it more likely that the will of the majority coincides with the voting outcome in the division of the assembly. It follows from the results by Curtis [16] that social welfare is maximized with a simple majority rule in the assembly and no voting blocs, but with simple majority, if there already exist some welfare-reducing voting blocs, new voting blocs may increase social welfare. For instance, imagine an assembly in which there is always a bare majority of agents in favor of the policy proposal, but there is one voting bloc that always rolls a few votes to the negative camp, swinging the outcome towards a rejection of the proposal. The creation of another bloc that rolls a few negative votes into favorable votes nullifies the negative effect of the first bloc and enhances social welfare.

Individual utility maximizing agents, however, are not concerned with social welfare or the effect of a bloc on the rest of society. They are only concerned from the benefit they derive from joining a voting bloc. Proposition 32 assures members of a bloc that collectively they benefit from its formation, but if agents cannot make compensating transfers, a surplus for a coalition does not guarantee a benefit to each of its members, and even if they all benefit, some agents may still have an incentive to leave and receive the benefits of the bloc as an externality without bearing the costs. The main goal in this subsection is to find stable voting bloc structures in the assembly when any arbitrary coalition of agents can form a voting bloc. I consider alternative definitions of stability.

The first notion is the already familiar Individual-Exit stability, which only requires voluntary participation in voting blocs, so that each agent is free to leave and become an independent. I now define the concept more rigorously for a voting bloc structure.

Definition 14 *A voting bloc structure* (π, r) *is Individual-Exit stable if* $u_i(\pi, r) \geq u_i(\pi', r)$ *for any* $i \in \mathcal{N}$ *and any partition* $\pi' = \{C'_j\}_{j=0}^{J}$ *such that:*

(i) $l \in C'_j \iff l \in C_j$ *for all* $l \in \mathcal{N} \backslash i$ *and all* $j \in \{0, 1, ..., J\}$, *and*

(ii) $i \in C'_0$.

Informally, a voting bloc structure is Individual-Exit stable if each of its voting blocs is itself Individual-Exit stable. This stability concept is similar, but less restrictive than the Individual Stability used by Drèze and Greenberg [19].

Definition 15 *A voting bloc structure* (π, r) *is Individually stable if* $u_i(\pi, r) \geq u_i(\pi', r)$ *for any*

$i \in \mathcal{N}$ and any partition $\pi' = \{C'_j\}_{j=0}^{J}$ such that:

(i) $l \in C'_j \iff l \in C_j$ for all $l \in \mathcal{N} \backslash i$ and all $j \in \{0, 1, ..., J\}$, and

(ii) for $j \in \{1, ..., J\}$, if $i \in C'_j$ then $u_l(\pi', r) \geq u_l(\pi, r)$ for all $l \in C'_j$.

Individual-Exit stability considers deviations only by departure from a bloc; Individual Stability allows for entry if it benefits every member of the coalition that receives an entrant. Entry is even more fluid under Nash stability; in a Nash stable voting bloc structure each agent is free to leave its bloc to become an independent or to migrate to any other bloc. In a Nash stable voting bloc structure every agent belongs to the bloc she likes most.

Definition 16 *A voting bloc structure (π, r) is Nash stable if $u_i(\pi, r) \geq u_i(\pi', r)$ for any $i \in \mathcal{N}$ and any partition $\pi' = \{C'_j\}_{j=0}^{J}$ such that $l \in C'_j \iff l \in C_j$ for all $l \in \mathcal{N} \backslash i$ and all $j \in \{0, 1, ..., J\}$.*

It follows from the definitions that the set of voting bloc structures that are Nash stable is contained in the set that are Individually Stable, which is itself contained in the set of Individual-Exit stable voting bloc structures.

Once stable voting bloc structures are identified, it is important to know if they have any effect in the outcome. Taking the structure with no voting blocs in which all agents act as independents as a benchmark, I analyze whether or not the formation of voting blocs affects the policy outcomes at least under some preference profile. If it never affects the policy outcomes, the coordination of voting behavior prompted by the voting blocs is irrelevant.

Definition 17 *Let (π^0, r^0) be the voting bloc structure in which all agents remain independent. A voting bloc structure (π, r) is relevant if with positive probability the policy outcome under the structure (π, r) differs from the outcome under (π^0, r^0).*

In short, if there is a relevant stable voting bloc structure, the coordination of voting behavior inside the blocs affects the policy outcome. It is possible to apply a similar definition to specific voting blocs, rather than to the whole structure. A particular voting bloc (C, r_C) is relevant if the coordination of votes inside the bloc (C, r_C) affects the policy outcome with positive probability. Formally:

Definition 18 *Let* (π, r) *be a voting bloc structure with* J *blocs* $j = \{0, 1, ..., \widehat{j}, ..., J\}$ *such that* $r_0 = 1$. *Let* (π, r') *be such that* $r'_{\widehat{j}} = 1$ *and* $r'_j = r_j$ *for all* $j \in \{0, ..., J\} \backslash \widehat{j}$. *The voting bloc* $V_{\widehat{j}} = \{C_{\widehat{j}}, r_{\widehat{j}}\}$ *is relevant in the structure* (π, r) *if with positive probability the outcome under the structure* (π, r) *differs from the outcome under* (π, r').

The next two results show that there exist relevant stable voting bloc structures.

Proposition 34 *Suppose* Ω *has full support and* $r_{\mathcal{N}} \leq \frac{N-1}{N}$. *Then there exists a relevant Individual-Exit stable voting bloc structure. In particular, any structure with a single voting bloc* (C, r_C) *such that* $N_C \geq r_{\mathcal{N}} N + 1$ *and* $r_C N_C < r_{\mathcal{N}} N$ *is relevant and Individual-Exit stable.*

A voting bloc that is more than large enough to act as a dictator is Individual-Exit stable because no agent gains anything by leaving a bloc that can still impose its will in the assembly after the defection. Since the outcome of the internal meeting of the bloc determines the outcome of the assembly, agents are better off participating in the internal meeting. The bloc is relevant because it needs a lower number of favorable votes to adopt the policy proposal -and impose it upon the assembly- than the threshold set by the voting rule of the whole assembly.

Proposition 35 *Suppose* Ω *has full support and* $r_{\mathcal{N}} \leq \frac{N-1}{N}$. *Then there exists a relevant Individually stable voting bloc structure. If* $r_{\mathcal{N}} \in (\frac{N+1}{2N}, \frac{N-1}{N}]$, *then a voting bloc structure with a single bloc* (C, r_C) *such that* $C = \mathcal{N}$ *and* $r_C < r_{\mathcal{N}}$ *is relevant and Individually stable.*

If the grand coalition forms a voting bloc, there is no possibility of deviating by entering a bloc. Hence the voting bloc structure is Individually stable -and Nash stable- if and only if it is Individual-Exit stable. From proposition 34, if the voting rule in the assembly is not simple majority, then a voting bloc by the grand coalition with a lower internal voting rule is relevant and Individual-Exit stable, hence it is Individually stable and Nash stable.

Corollary 36 *Suppose Ω has full support and $r_{\mathcal{N}} \in (\frac{N+1}{2N}, \frac{N-1}{N}]$. Then there exists a relevant Nash stable voting bloc structure. In particular, a voting bloc structure with a single voting bloc (C, r_C) s.t. $C = \mathcal{N}$, $r_C < r_{\mathcal{N}}$ is relevant and Nash stable.*

If the voting rule of the assembly is simple majority, then proposition 35 shows that a relevant Individually stable voting bloc structure exists. In particular, a voting bloc structure with a unique voting bloc (C, r_C) such that $N_C = N - 2$ and r_C is simple majority is relevant and Individually stable. The bloc acts as a dictator and its members do not want to leave and would not benefit by admitting any of the two non-members into the bloc. This particular voting bloc structure is not Nash stable because the two non-members, who are essentially excluded from the decision-making process, would enter the bloc is such a deviation was feasible for them. In fact, for some probability distributions over preference profiles, there is no relevant Nash stable structure. The next result shows existence of a Nash stable structure, and the following one describes characteristics of relevant Nash stable structures, provided they exist.

Proposition 37 *A Nash stable voting bloc structure exists.*

The grand coalition $C = \mathcal{N}$ with $r_C \in [r_{\mathcal{N}}, 1]$ is irrelevant, but Nash stable. In a bloc $(\mathcal{N}, r_C)$ with $r_C \geq r_{\mathcal{N}}$, if the majority in the bloc had enough supporters to roll i, then it has enough supporters to win in the division of the assembly, regardless of the rolled votes. Thus, an agent cannot change the outcome by leaving, and the bloc is Nash stable.

Proposition 37 relates closely to corollary 36: If the coalition of the whole forms a bloc with a lower internal voting rule than the rule used in the division of the assembly, the voting bloc is relevant. If it forms a bloc with a higher internal voting rule, it is irrelevant. In both cases it is Nash stable, but in the first one it effectively functions as if it endogenously changed the voting rule of the assembly, and in the second case it merely makes some proposals pass (or fail) with unanimity when they would have passed (or failed) just by majority.

If they exist, Nash stable voting blocs have to be of size smaller than minimal winning.

Proposition 38 *Suppose Ω has full support and $r_{\mathcal{N}} = \frac{N+1}{2N}$. Then in any relevant Nash stable voting bloc structure (π, r), $N_C < \frac{N+1}{2}$ for any voting bloc $\{C, r_C\}$ with a simple majority internal voting rule, and if there exist at least one singleton in (π, r), then $N_C < \frac{N+1}{2}$ for any relevant voting bloc $\{C, r_C\}$.*

Proposition 38 tells us that a relevant voting bloc cannot be large enough to act as a dictator. If it is, every agent would like to join. To illustrate this result, think of the solid Democratic South of the US during the first half of the 20th century. In an essentially one-party system, any politician with some aspirations of furthering his ideal policies through the State legislatures had a strong incentive to become a Democrat, irrespective of his political ideology. With no barriers to enter blocs, competition among opposing blocs only occurs if the weaker blocs also have a hope of influencing the policy outcomes.

None of the solution concepts that I have studied so far allows members of a bloc to coordinate a coalitional deviation. Hard as it may be for agents to communicate and coordinate across blocs, it seems easier to scheme a deviation in which a subset of members in a bloc defect together, and possibly form a new voting bloc. The following stability 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 19 *A voting bloc structure (π, r) with J blocs is Split stable if there exists no partition $\pi' = \{C'_j\}_{j=0}^{J+1}$, rule r_{J+1} and coalition $C_{\widehat{j}} \in \pi$ such that:*

(i) For all $j \in \{1, ..., J\}$, $r'_j = r_j$,

(ii) for all $j \in \{1, ..., J\} \backslash \widehat{j}$ and all $i \in \mathcal{N}$, $i \in C'_j \iff i \in C_j$,

(iii) for all $i \in \mathcal{N}$, $i \notin C_{\widehat{j}} \implies i \notin C'_{\widehat{j}}$,

(iv) for any $i \in \mathcal{N}$ s.t. $i \in C'_0$ and $i \notin C_0$, $u_i(\pi', r \cup r_{J+1}) \geq u_i(\pi, r)$ and for any i s.t. $i \in C'_{J+1}$, $u_i(\pi', r \cup r_{J+1}) > u_i(\pi, r)$.

Definition 20 *A voting bloc structure* (π, r) *is weakly Split Stable if there exists no partition* $\pi' = \{C'_j\}_{j=0}^{J+1}$ *and bloc* $V_{\widehat{j}}$ *satisfying conditions (i), (ii) and (iii) in Definition 19 and*

(iv)' For any $i \in \mathcal{N}$ *and* $j \in \{0, \widehat{j}\}$*, if* $i \in C_j$ *and* $i \notin C'_j$ *then* $u_i(\pi', r \cup r_{J+1}) > u_i(\pi, r)$.

Conditions *(i)* and *(ii)* say that all other blocs remain unaffected by the coalitional deviation involving independent agents and members of bloc $V_{\widehat{j}}$; both the internal rules and the membership of these blocs remain the same, and the internal rule $r_{\widehat{j}}$ of the bloc $V_{\widehat{j}}$ is also kept intact. Condition *(iii)* states that in the new partition the bloc that suffered the defection gains no new members. 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 would 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 slightly stricter condition *(iv)'* foregoes this refinement over indifference, and accepts as weakly Split stable any voting bloc structure in which agents do not have a strictly positive incentive to deviate.

The intuition for the Split notion of stability[5] 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 (i.e., Western and Soviet blocs during the Cold War); whereas, a disaffected subset of a bloc can more easily break apart and possibly recruit some independent agents for a new voting bloc. As an example, the moderate wing of the UK's Labour party broke off in 1981 and formed the Social Democratic Party, which attracted up to 28 former Labour MPs.[6]

The notion that some members of a coalition may organize a coordinated defection even though deviations across coalitions are not feasible is common to two previous concepts of equilibrium in the non-cooperative coalition formation literature: The Coalition-Proof equilibrium by Bernheim, Peleg

[5]Kaminski [34] and [35] first introduced the term Split Stability to denote a similar but different concept in which coalitions of parties can split, in an applied study of political parties in Poland.

[6]Admittedly, cross-party deviations are sometimes also successful, as illustrated by the new Kadima party in the Israeli Knesset.

and Whinston [8] and the Equilibrium Binding Agreements by Ray and Vohra [48]. In these two concepts, agents negotiate as if each coalition was in a separate room, and any group of agents in the same room could leave and find a new room for themselves, with the important proviso that every deviation must itself be immune to further deviations (once they deviants reach their new room, it must be that no subset of them would want to leave for yet another room), so the definitions are recursive.

Split stability is different first in that it is not a recursive concept, since I do not require a coalition of deviants to be immune to further deviations. Second, while I do not consider deviations across coalitions, I allow deviants to coordinate with independents. Under Split stability, 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.

Why use Split stability? First, although the requirement that deviations be themselves immune to further deviations adds consistency to the Coalition-Proof and Equilibrium Binding Agreements solution concepts and makes them theoretically more elegant, it also adds a layer of unwelcome complexity. Allowing for best-response type of deviations makes Split stability a much simpler concept to define and use in applications. Second, agents may prefer to proceed with a deviation even if this deviation is itself unstable. If it is difficult to predict the stream of deviations that will follow an initial departure and agents cannot anticipate the ultimate outcome propitiated by their deviation, a subset of agents displeased with the original configuration may choose to deviate for a short-sighted gain even if subsequent deviations by other agents could potentially undo the improvement brought by their first move. If so, Coalition-Proof and Equilibrium Binding Agreements solutions will be unstable.

Previous literature has shown that existence of stable coalition structures becomes problematic as soon as coalitional deviations are allowed. Hart and Kurz [29] consider five alternative notions of stability with coalitional deviations, including core stability and the Strong Nash equilibrium of an associated membership game called the Δ game in which agents simultaneously choose which coalition to join. Hart and Kurz [30] then show that even the largest of these five solution sets may be empty. Split stability restricts the possible coalitional deviations to those involving members of at most one voting bloc. As a result, the solution set is larger than core stability or Strong Nash equilibrium of the Δ game. Under some restrictions on the probability distribution of preference profiles, Split stable voting bloc structures exist.

Proposition 39 *Suppose types are independent, $t_i = t \in (0,1)$ for any $i \in \mathcal{N}$ and $r_{\mathcal{N}}$ is simple majority. Then there exists a relevant weakly Split stable voting bloc structure (π, r) in which as many voting blocs of size three as feasible form and there remain zero, one or two singletons, depending on N. If N is not a multiple of three, this relevant voting bloc structure is Split stable.*

A complete characterization of Nash or Split stable voting bloc structures in full generality is an overly ambitious task, since the solution varies with the size of the assembly and the probability distribution over preference profiles of its members. Instead, in the following section I illustrate the model at work applying the theory to a small assembly of size nine, using the empirical data from the United States Supreme Court.

4.4 A Small Assembly: The US Supreme Court

The theoretical model of endogenous voting blocs at the end of the previous section showed that there exist stable partitions of an assembly into voting blocs, but besides predicting that Nash stable voting blocs are of size less than minimal winning, it provided limited information about the features of these stable voting bloc structures. In this section I seek a more detailed description of the voting blocs that we expect to find in a small assembly. First I provide the theoretical prediction in a stylized assembly of size nine, and then I look at data from the United States Supreme Court from 1995$-$2004.

4.4.1 Endogenous Voting Blocs in a Small Assembly

Consider an assembly with nine voters who have independent and symmetric types distributed as follows: t_1=t_2=$0.5 - \alpha - \beta$; t_3=t_4=$0.5 - \alpha$; t_5=0.5; t_6=t_7=$0.5+\alpha$; t_8=t_9=$0.5+\alpha+\beta$, with $\alpha, \beta \geq 0$ and $\alpha + \beta \leq 0.5$. That is, types 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 types of length 2α. The more polarized the members of the assembly are, the larger α and the larger the differences in types 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.

Using numerical simulation for a fine grid of values of α and β, I find which connected voting bloc structures are relevant and Nash and Split stable such that no agent or group of agents would have an incentive to deviate to a different structure.

Definition 21 *A voting bloc structure (π, r) is* ***connected*** *with respect to the order $<$ if for all $C \in \pi$ and for all $i, j, k \in \mathcal{N}$, $(i, k \in C$ and $i < j < k)$ implies $j \in C$.*

The order of agents is according to their type, so a voting bloc is connected if its members are in consecutive positions in the ordering by types. Axelrod [4] provides a detailed argument in favor of connected coalitions over non-connected ones.

The intuition that in a very polarized assembly there will not be a unique moderate bloc, but rather, two blocs one on each side of the political spectrum is verified. There are only four connected voting bloc structures that are Nash and Split stable. These are all such that exactly two blocs L and

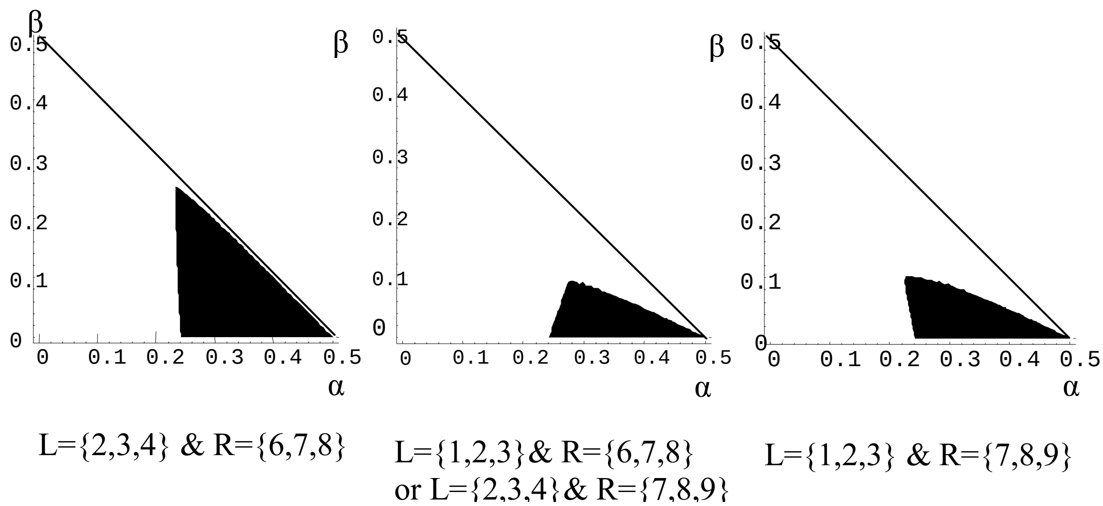

Figure 4-2: a,b,c. Nash and Split stable voting bloc structures.

R form with simple majority internal voting rules, each with three members, and $L \subset \{1, 2, 3, 4\}$, $R \subset \{6, 7, 8, 9\}$. That is, three of the four members of the assembly with a low type form a voting bloc, and three members with a high type form another bloc. It is easy to visualize the L bloc as a pro status quo party, which tends to vote against the policy proposal, and the R bloc as a reform party, which tends to vote for the policy proposal.

Figure 4-2 a,b,c shows in black the parameter values for which each voting bloc structure is Nash and Split stable. For any $\alpha < 0.5$, the voting bloc structure is relevant.

The three figures share the common characteristic that only for a high α the voting bloc structure is stable. If the assembly is not polarized and agents share similar types, then each agent in voting bloc L has an incentive to defect to R, effectively disbanding L since no bloc can function with only two members. If there is enough polarization, defections across blocs no longer occur.

As a summary, this subsection has shown that if the assembly is sufficiently polarized, there is a stable and relevant connected voting bloc structure composed of two opposing blocs, located one at each side of the median. In the rest of the section I depart from the stylized assumptions of the modelled assembly (symmetry and independence of types), looking instead at real data from the United States Supreme Court. After introducing the Court and the policy preferences of its members, I calculate the effect of voting blocs upon the outcome of the Court.

4.4.2 The United States Supreme Court

The United States Supreme Court is the ultimate appellate court in the United States judicial system, and the arbiter of the United States Constitution. It is composed of nine justices and it uses a simple majority rule, so that the vote of five justices are enough to decide a case. The Court makes a binary decision on the merits of each case: It either affirms or reverses the ruling of a lower court. In an accompanying *opinion,* the Court provides the argumentation for its decision, and this *opinion* serves as precedent for future cases.

I use the data on the decisions of the Court from *The United States Supreme Court Judicial Database* compiled by Spaeth [58] and I select all non-unanimous cases with written opinions in which all nine justices participate.[7] Spaeth codes the votes of each justice 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. An alternative binary coding of the votes which is unambiguously objective divides the votes between votes with the majority, and dissents -votes with the minority.

Table 4.2 shows the number of liberal votes and the number of dissents that each justice cast in the 419 non-unanimous decisions recorded from 1995 and 2004. The nine justices, abbreviated by the first three letters of their surname, are: Stevens, Ginsburg, Souter, Breyer, O'Connor, Kennedy, Rehnquist, Scalia and Thomas.

Table 4.2: Liberal and dissenting votes in 419 decisions

	1.Ste	2.Gin	3.Sou	4.Bre	5.O'Co	6.Ken	7.Reh	8.Sca	9.Tho
Liberal	344	308	307	276	160	155	98	84	71
Dissent	203	159	136	141	71	78	115	161	156

[7]The unit of analysis in my data is the case citation (ANALU=0), the type of decision (DEC_TYPE) equals 1 (orally argued cases with signed opinions), 6 (orally argued *per curiam* cases) or 7 (judgments of the court), and I drop all unanimous cases and all cases in which less than 9 justices participate in the decision.

The most extreme justices, either liberal or conservative, find themselves in the minority of dissenters more often than the moderate justices. Justice O'Connor, traditionally regarded as the swing justice, dissents in only about one in six cases, while Justice Stevens, who is the most liberal member of the Court, dissents from the majority in roughly a half of the cases.

If justices formed voting blocs, the coordination of votes would change the voting record of the justices, the composition of the majority and dissent justices in each case, the outcome of some decisions, and, assuming that justices are policy oriented, the utility or satisfaction of the justices with the outcome of the Court. I calculate the changes brought by any possible connected voting bloc structure in the Court.

The notion of a connected voting bloc requires an ordering of justices from one to nine. In the tables and the text of this section I use the ordering according to the number of liberal votes cast as recorded by Spaeth [58]. I check if this ordering is robust by means of calculating the ideal location of the justices in a space vector using three mathematical methods that abstract from the substantive content of each case and attend only to the voting patterns and correlations across the justices. Although my basic goal is to obtain an objective and robust ordering of the justices, these analyses have an intrinsic value in that they provide estimates of the location of each justice in a vector space with an easy interpretation in ideological terms such as a liberal/conservative scale.

The three methods I use are: Singular Value Decomposition of the original data, Eigen Decomposition of the square matrix of cross-products of the locations of the justices, and the Optimal Classification method developed by Poole [45], and I compare these three estimates with the findings of Martin and Quinn [40] and [25], who use Bayesian inference in a probabilistic voting to estimate the ideal points of the justices.[8] In Table 4.3 I provide the ideal position of the justices estimated by Single Value Decomposition and Eigen Decomposition, the rank ordering given by the Optimal Classification method in one dimension, the estimate of the position in the first dimension given by

[8]See Brazill and Grofman [11] for a comparison of the relative merits of multidimensional scaling methods versus factor analysis methods such as the Eigen or Single Value Decomposition.

the Optimal Classification method in two dimensions, and the estimates obtained by Martin and Quinn. First I briefly explain each of the methods.

Mathematically, the Singular Value Decomposition of a rectangular matrix $\mathbf{X}_{419*9}$ is

$$\mathbf{X}_{419*9} = \mathbf{U}_{419*419}\mathbf{D}_{419*9}\mathbf{V}_{9*9} \text{ s.t. } \mathbf{U}^t\mathbf{U} = \mathbf{I} \text{ and } \mathbf{V}^t\mathbf{V} = \mathbf{I}.$$

The matrix $\mathbf{X}$ contains the original data of zeroes (dissents) and ones (votes with the majority), each case in a row and each justice in a column. This original data is decomposed into two orthogonal matrixes and a diagonal matrix. The vectors in the square matrix $\mathbf{V}$ represent the estimates of the ideal point of each justice in nine new dimensions, such that the estimates for the first dimension represents the best fit to the original data with only one dimension; the estimates for the second dimension are the best fit adding a second dimension but taking the estimates for the first dimension as given, and the estimates for the kth dimension are the best fit in k dimensions taking the previous $k-1$ dimensions as given. Here, "best fit" means the approximation that minimizes the sum of the squared error between the approximation and the original data. The "single values" along the diagonal of $\mathbf{D}$ are all positive and represent the weights of each of the dimensions. See Eckart and Young [20] for the original mathematical idea.

The Single Value Decomposition generates nine new coordinates capturing the most frequent alignments of voting in the Court, and gives the location of each justice in all nine dimensions, so that taking only the first one or two dimensions gives the best approximation of the location of the justices in this reduced subspace.

The Eigen Decomposition and the Optimal Classification method require some previous steps. First, I calculate the disagreement matrix, which is a 9 by 9 matrix that shows for each pair of justices, the proportion of cases in which they do not vote together. Second, I convert the disagreement score matrix into a matrix of squared distances, just by squaring each cell. Third, I double center the squared distances matrix by subtracting from each cell the row mean and the column mean, adding the matrix mean, and dividing by -2. Double centering the squared distances matrix removes the squared terms and produces a cross-product matrix of the legislator coordinates. For details of these steps, see Poole [45]. The Eigen Decomposition of the cross-products matrix produces nine eigenvectors, which we can interpret as estimates of the location of the justices in nine dimensions, and nine corresponding eigenvalues, which assign weights to each of the dimensions. Mathematically, the Eigen Decomposition of a square matrix $\mathbf{X}_{9*9}$ is

$$\mathbf{X}_{9*9} = \mathbf{U}_{9*9}\ \mathbf{D}_{9*9}\ \mathbf{U}_{9*9}^{-1},$$

where the elements of the diagonal are the eigenvalues, and the vectors of $\mathbf{U}$ the eigenvectors.

The Optimal Classification method in one dimension applied to the Supreme Court data ranks justices from one to nine, and ranks each case in between a pair of justices, predicting that all justices to one side will vote one way, and all justices on the other side will vote the other way. For instance, if a case is ranked between 2 and 3, the OC method predicts that justices 1 and 2 vote in the minority and the other seven justices in the majority. If in the real data justice 3 also votes with 1 and 2, then that is one classification error, and the OC method aims to minimize the number of these errors.

The algorithm used in the Optimal Classification method is as follows. Starting with the rank ordering of the justices given by the first vector of the Eigen Decomposition of the double-centered squared-distances matrix, assign a rank to every case in such a way that the ranks minimize the total number of errors. Then, given the rank of every case, assign a new rank to the justices to minimize the number of errors given the ranking of cases. The algorithm proceeds iteratively reranking cases given the ranking of justices and then reranking justices given the ranking of cases until it converges to a solution that jointly gives a rank of both justices and cases that minimizes the number of classification errors. In two dimensions, instead of rank orderings, the OC method assigns a position in the space for each justice -or, more precisely, an area where the justice is located- and for each case it gives a cutting line partitioning the space into the area where it predicts that justices vote with the majority and the area where it predicts that justices vote with the minority. Poole [45] provides a careful explanation of this method.

To my best knowledge, the most complete analysis of the location of the ideal policies of recent Supreme Court justices is the Supreme Court Ideal Point Research conducted by Martin and Quinn [40] and [25], who use a probabilistic voting model and Bayesian inference to estimate the ideal policies of the justices in a unidimensional space. A particularly useful feature of their project is that they study the dynamics of the Court, and they update they results year by year at the homesite of the project at adw.wustl.edu/supct.php. I take the average of the estimates they report for the years 1995−2004. Estimates by Singular Value Decomposition and the Optimal Classification method range from minus one (most liberal) to plus one (most conservative). Martin and Quinn's estimates could take any value in the real line, but since the scaling of their estimates is arbitrary, I rescale their estimates dividing by five to ease the comparison across rows in Table 4.3.

Table 4.3: Estimates of the location of ideal policies

	1.Ste	2.Gin	Sou	Bre	5.O'Co	6.Ken	7.Reh	Sca	Tho
SVD	−0.425	−0.382	−0.351	−0.335	0.089	0.154	0.294	0.398	0.459
Eigen D	−0.418	−0.296	−0.250	−0.253	0.161	0.212	0.348	0.455	0.459
OCM 1D	$1st$	$2nd$	$4th$	$3rd$	$5th$	$6th$	$7th$	$8th$	$9th$
OCM 2D	−0.736	−0.583	−0.506	−0.498	0.169	0.274	0.489	0.704	0.661
M&Q	−0.590	−0.302	−0.248	−0.221	0.099	0.146	0.289	0.598	0.678

As shown in the table, the different methods produce similar estimates that mostly corroborate the initial ordering of the justices according to the proportion of liberal votes cast, as coded by Spaeth [58].

The ordering according to Martin and Quinn and according to the Single Value Decomposition (SVD) coincides exactly with the ordering according to the proportion of liberal votes. It is important to note that the estimates from the SVD in the table correspond to the second dimension of the SVD. The first dimension is an "agreement dimension" in which all justices get a very similar value; this dimension captures the insight that justices tend to vote together very frequently, and it is only the second vector that provides the relevant information of the location of the justices in the dimension of interest. I report the estimates for the first dimension and the weight for all nine dimensions in the appendix. Sirovich [56] used the same method to study the voting patterns of the Court from 1995 to 2002, and his estimates are similar to mine as was to be expected, with two differences. First, he fails to omit the unanimous decisions. As a consequence, the first dimension in his analysis is more accurately an agreement dimension in which all justices get an approximately equal estimate, and this (uninteresting) "agreement dimension" carries more weight than in my analysis. Second, in my data Justice Souter appears to be more liberal. This reflects the fact that Justice Souter gradually drifted during his tenure in the Court, a fact also recorded by Martin and Quinn.

The estimates according to the first eigenvector of the Eigen Decomposition of the cross-product of justices' coordinates switch the positions of Souter and Breyer by a very slim margin, and otherwise coincide with the proportion of liberal votes or the estimates by SVD.

The Optimal Classification method with one dimension again switches the ordering of Souter and Breyer, but with two dimensions, Optimal Classification returns Souter back to the left of Breyer and it alters the ordering of Scalia and Thomas.

All estimates agree in the following partial order $\prec$:

$$Ste \prec Gin \prec \begin{matrix} Sou \\ Bre \end{matrix} \prec O'Co \prec Ken \prec Reh \prec \begin{matrix} Sca \\ Tho \end{matrix}.$$

The only open questions are the relative ordering of Breyer and Souter, and the relative ordering of Scalia and Thomas. Rather than making a questionable assumption about these two pairs of justices, I consider all four lineal orders consistent with the partial order $\prec$ and I evaluate all the voting bloc structures that are connected according to one of these four lineal orders. Formally, a partial order is a binary relation that is reflexive, transitive and antisymmetric. A lineal order adds the property of being total, that is, it orders every pair of elements. For instance, the coalition $C = \{Ste, Gin, Sou\}$ is connected given the partial order $\prec$ because it is connected given the linear order that ranks Souter third and Breyer fourth and the coalition $C' = \{Ste, Gin, Bre\}$ is also connected given $\prec$ because it is connected given the linear order that ranks Breyer third and Souter fourth. But if a coalition contains both Gin and $O'Co$, then it must contain both Bre and Sou to be connected given $\prec$.

4.4.3 Endogenous Voting Blocs in the US Supreme Court

> "People ask me whether I was sorry that I was in the minority in Bush vs Gore. 'Of course I was sorry!' I'm always sorry when I don't have a majority." Justice Stephen Breyer of the US Supreme Court.[9]

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, as quoted from Justice Breyer, wants the Court to reach a decision according to the preference of the justice. This assumption is consistent with the attitudinal model of the Court by Segal and Spaeth [55], who consider competing models of the functioning of the Court and conclude that a model of sincere voting by policy-oriented justices best explains the decisions of the Court. In earlier work, Rohde [52], studied the formation of coalitions in the writing of opinions in the Warren Court (1953−1968) and assumed that the optimization problem of the justices is to have the policy output of the Court approximate as closely as possible his own preference. If Segal and Spaeth [55] are correct and justices vote sincerely, then each justice wanted the decision of the Court to coincide exactly with the vote that the justice cast and every dissent is a defeat. Even if justices do not always vote sincerely, it would be difficult to discern the true preferences of the justices beyond their revealed preferences, so I assign utilities according to the actual votes cast by the justices.

[9] From "Breyer's Big Idea," by Jeffrey Toobin, in *The New Yorker*, October 31st, 2005, pages 36−43.

The Court makes a binary decision on the merits of each case: It either affirms the ruling from a lower court, or it reverses it; it sides with the plaintiff, or with the defendant; with the liberal position, or the conservative one. For instance, in a case in which a lower court took a conservative view and sided with the plaintiff, the outcome of the decision is either affirm-plaintiff-conservative or reverse-defendant-liberal. 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 justices, as revealed by the vote of the justice. Table 4.2 then provides the ultimate satisfaction of each justice with the series of decisions of the Court: 419 minus the number of dissents is my measure of the utility or satisfaction of each justice with the output of the Court from 1995 to 2004. This measure of utility implicitly assumes that justices only care about how often they obtain a majority, or in other words, that they do not care more about some decisions over others. While this assumption is admittedly unrealistic, it is a simplifying step to circumvent the need to assign weights for each case and justice.

I calculate how the outcomes would have changed if justices had formed voting blocs, and how the satisfaction of each justice would have changed accordingly. For a given voting bloc structure in the Court, I assume that each bloc holds a private internal vote before the division of the Court, and in these internal votes I assume 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 bloc structure 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.

Example 7 *Suppose Ginsburg, Souter and Breyer form a voting bloc. Then the net change in the number of decisions in which each justice is satisfied with the outcome is as follows:*

Bloc	*1.Ste*	*2.Gin*	*3.Sou*	*4.Bre*	*5.O'Co*	*6.Ken*	*7.Reh*	*8.Sca*	*9.Tho*
{234}	*12*	*12*	*4*	*4*	*-2*	*-14*	*-10*	*-12*	*-14*

Example 7 shows that had Ginsburg, Souter and Breyer committed to always vote together rolling internal dissent among the three, each of them would have achieved their preferred outcome more often even if sometimes they had to vote against their preference. Comparing these numbers to those in Table 4.2, Ginsburg would reduce the number of cases that end up against her preference by almost 8%. Souter and Breyer by about 3%.

If justices Ginsburg, Souter and Breyer had formed a voting bloc, 20 decisions out of 419 would have been reversed, *Atwater vs City of Lago Vista* (2001) among them. In a 5−4 decision, the Court held that the Fourth Amendment does not forbid a warrantless arrest for a minor criminal offense, such as a misdemeanor seatbelt violation punishable only by a fine. Justices Souter, Kennedy, Rehnquist, Scalia and Thomas voted with the majority. Justice O'Connor, joined by Stevens, Ginsburg and Breyer, wrote a dissent arguing that a seatbelt violation is not a reasonable ground for arrest, and thus the arrest is in violation of the Fourth Amendment that prohibits unreasonable seizure. With the exception of Souters, there is a clean division of the Court between more liberal justices favoring broader Civil Rights, and more conservative justices favoring Law Enforcement. Had Souters voted with Ginsburg and Breyer, the Court would have found the arrest to be unconstitutional.

More recently, in two famous cases decided on June 27, 2005, the Court ruled that the display of the Ten Commandments in two courthouses in Kentucky is in violation of the First Amendment Establishment Clause for the Separation of Church and State, but it also ruled that a display of the Ten Commandments in the Texas State Capitol is not unconstitutional. Justices Stevens, Ginsburg, Souter and O'Connor voted against the displays both in the Kentucky and Texas cases, while justices Kennedy, Rehnquist, Scalia and Thomas voted in favor of the displays in both cases. Justice Breyer voted against the Kentucky displays in *Mc Creary County vs ACLU*, giving the liberals a 5−4 majority, but he voted in favor of the Texas display in *Van Orden vs Perry*, giving the conservatives a $5-4$ majority. Had Breyer voted with Souter and Ginsburg in both cases, the Texas display would have been ruled unconstitutional, just as the Kentucky ones.

Note that when a justice in a voting bloc has to vote against his true preference in the division of the Court, he would only be satisfied with the outcome if his vote -along with the whole bloc he belongs to- ends up in the minority of the Court. Hence Example 7 does not measure the extra number of times that Ginsburg, Souter or Breyer are in the majority, but the extra number of times that they are satisfied with the outcome. In so far as justices are ideologically motivated, it is reasonable to say that for a justice *to win* means that the preferred outcome of this justice prevails, regardless of whether the justice voted for or against her favored outcome in the division of the Court.

Epstein and Knight [24] argue that justices make strategic choices deviating from their preference for the sake of achieving the policy outcomes they desire so that the Law that emanates from the Supreme Court rulings is "the long term product of short-term strategic decision-making." I argue that if justices are strategic in their actions, then they must be tempted to devise not just short-sighted strategies for one case, but rather long-term strategic plans such as forming a voting bloc. For instance, if Justice Breyer had formed a voting bloc with Ginsburg and Souter and no other justice had reacted to that bloc, Justice Breyer would have lost fewer cases, exactly four less.

Assume the counterfactual that Ginsburg, Souter and Breyer form a voting bloc. This bloc is Individual-Exit stable because all three members benefit from joining so none would want to deviate and leave disbanding the bloc. However, the voting bloc structure in which Ginsburg, Souter and Breyer form the only voting bloc is neither Nash stable, nor Split stable, because other justices have incentives to react to this bloc. Table 4.4 displays the net payoffs to each justice relative to the benchmark with no voting blocs if Stevens joins the bloc (first row), and if the Rehnquist, Scalia and Thomas form another voting bloc (second row). A summary comparison between Table 4.4 and the table in Example 7 reveals that Stevens would benefit if he joined the liberal bloc, increasing his net utility from +12 to +13. Hence the voting bloc $Gin - Sou - Bre$ is not Nash stable. The second row reveals that Rehnquist, Scalia and Thomas would reduce their loses from the formation of the $Gin - Sou - Bre$ bloc if they formed their own bloc.

Table 4.4: Net change in satisfaction

Blocs	1.Ste	2.Gin	3.Sou	4.Bre	5.O'Co	6.Ken	7.Reh	8.Sca	9.Tho
{1234}	13	11	9	5	1	−15	−17	−13	−13
{234},{789}	5	−5	−7	−3	9	−1	−7	−3	−9
{123},{789}	−1	−5	−5	3	9	−3	−7	−1	−11
{789}	−5	−9	−13	−9	1	1	1	13	3
{123}	8	6	2	8	0	−10	−12	−6	−12

A single bloc with the four liberal justices is Nash stable, but it is not Split stable, because Rehnquist, Scalia and Thomas have an incentive to form their own bloc to counterbalance the four liberals just as much as they do against a bloc of three liberals.

On the other hand, the voting bloc structure with both a liberal bloc formed by Ginsburg, Souter, Breyer and a conservative bloc formed by Rehnquist, Scalia, Thomas is Nash stable and Split stable. With the partial ordering $\prec$ discussed above, the following result summarizes my findings on the stability of hypothetical voting blocs in a nine agent assembly whose members faced an agenda and preference profile identical to those of the US Supreme Court justices between 1995−2004.

Result 40 *Any voting bloc structure in which 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 is Nash and Split stable, and there exists no other Split stable connected voting bloc structure.*

I show the payoffs -net changes in the number of cases that each justice wins- given that $\{Ste, Gin, Sou\}$ form a voting bloc, and the three most conservative justices form another voting bloc in row three of Table 4.4. For the sake of comparison, in rows four and five I show the payoffs if one member of the liberal bloc deviates and the bloc dissolves, leaving $\{Reh, Sca, Tho\}$ as the unique bloc, and the payoffs if the conservative bloc dissolves and $\{Ste, Gin, Sou\}$ remain as a bloc. It is clear that no justice wants to abandon the bloc he belongs to.

In the appendix I provide a table with the payoffs for each justice for a sample of voting bloc structures with a single bloc and for the two Split Stable voting bloc structures not already listed in Table 4.4. For all other connected voting bloc structures, I provide the payoffs for each justice and a deviation (if any exists) that makes such structure not Nash stable or not Split stable in an Excel file that also contains the original data and formulas to replicate the calculations. This file is available at www.hss.caltech.edu/~jon, or directly from the author.

According to Result 40, the stable voting bloc partitions 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. Stable voting bloc partitions merely reinforce the polarization of the Court into a liberal group and a conservative group of justices, and do not produce a major realignment of votes. As a curiosity, the most famous decision of this court, the $5-4$ division in Bush vs Gore (2000) which stopped the recount of the Florida votes and gave Bush the presidency would not have been reversed, since it was already the case that the four liberals voted together in the minority, and the two moderate conservatives and three conservatives voted together in the majority. The formation of a stable connected voting bloc would have made this particular $5-4$ conservative-liberal division more frequent, but this was already the most frequent split of the Court.

Consider the stable voting bloc structure in which Stevens, Ginsburg and Souter form a voting bloc, and Rehnquist, Scalia and Thomas form another voting bloc. In terms of the location of each of the blocs in an ideological space, each of the blocs converges near the location of its median member. Indeed, the liberal bloc casts 313 liberal votes where Ginsburg alone casts 308 (Stevens and Souter cast 344 and 307, see Table 4.2) and its position in space according to SVD is -0.318, indistinguishable from Ginsburg's in the dimension of interest, while the conservative bloc casts 72 liberal votes for 84 of Scalia (98 and 71 by Rehnquist and Thomas) and locates by SVD at 0.392, where Scalia alone is at 0.398.

Compare Result 40 with the stylized assembly with 9 agents who have symmetric and independent types. Note that the stable voting bloc structures in Result 40 are a subset of those in the idealized assembly. The theory predicted that stable voting bloc structures would consist of two 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 this set of stable voting bloc structures, and the only difference is that the conservative bloc has to be $\{789\}$ and cannot be $\{678\}$ instead. The cause of this difference is that the modelled assembly assumed that agents 3 and 4 and agents 6 and 7 are identical. In the empirical application, Souter and Breyer are indeed similar enough in their voting behavior, but Kennedy is markedly different from Rehnquist, and in particular Kennedy is not conservative enough to benefit from forming a bloc with Rehnquist and Scalia.

The following two comments suggest that Result 40 should be interpreted with caution.

First, the Split stable voting bloc structures are in accordance to those predicted by the more abstract model and reinforce the intuition that the assembly is likely to split into two opposing voting blocs that counterbalance each other, one at each side of the ideological spectrum and leaving a number of unaffiliated moderate independents. However, this result is based on the particular stability concept that I have chosen. The question of which equilibrium refinement or which stability concept is appropriate is still open in the literature.

Second, this section has shown what voting bloc structures would be stable 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. I leave to Supreme Court scholars these tasks. Restraints of a legal, normative or ethical nature may deter Supreme Court justices from committing to vote as a bloc, and this section does not attempt to explain voting in the US Supreme Court as much as it intends to illustrate how voting blocs could affect outcomes in practice, and what voting bloc structures would be stable. The data from the US Supreme Court serves by proxy to shed some light into the formation of voting blocs in committees, councils, small assemblies, and all sorts of political caucuses, in which the incentives to form the blocs will be salient and the restraints that Supreme Court justices face are probably absent -and, crucially, the data on the preferences of its members is also absent.

I have proved that 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 chapter.

4.5 Conclusion and Extensions

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 stable voting bloc structures exist for various concepts of stability in a model in which agents with heterogeneous preferences coalesce into voting blocs endogenously.

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.

I have illustrated how voting blocs affect voting outcomes using data from the US Supreme Court decisions between 1995 and 2004.

The theory in this thesis has multiple natural extensions: Comparing the results under other stability concepts, such as Coalition-Proofness or Equilibrium Binding Agreements; endogenizing the choice of the internal voting rule for each bloc and allowing for a richer class of rules, not just anonymous and majoritarian rules; studying the enforceability of the internal rules in a repeated game if binding commitments are not feasible; introducing intensity of preferences so that agents who like the proposal do so to varying degrees; considering unequally weighted individuals or even pyramidal structures, in which individual agents coalesce into factions, factions coalesce into parties (voting blocs of second order), parties into alliances (voting blocs of third order) and so on. Empirical applications range from revisiting the historical records of the early United States Congress to try to determine the incentives to coordinate votes along state lines or along parties, to salient current developments such as the theoretical advantages to each of the 27 European Union countries from pooling their votes under a common foreign EU policy. These questions constitute an agenda for further research.

4.6 Appendix

4.6.1 Proof of Lemma 25

To prove this lemma I use Claim 23 from the Appendix to chapter 3, and another intermediate result, which shows that if M is symmetric and L leans left and forms a voting bloc, then the distribution of the number of votes cast by $L \cup M$ in the division of the assembly is such that given any absolute difference between the number of votes $L \cup M$ casts for and against the proposal, the net difference is negative with probability at least a half.

Claim 41 *Let $\mathcal{N} = L \cup M \cup R$. Suppose $g^M(\frac{N_M-1}{2} - k) = g^M(\frac{N_M+1}{2} + k)$, $g^L(\frac{N_L-1}{2} - k) \leq g^L(\frac{N_L+1}{2} + k)$ for all non-negative k and L forms a voting bloc (L, r_L). Then:*

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

Proof. Lets first define L to be *active* given a preference profile s if it rolls its internal minority given the rule r_L, so that $s_i \neq v_i$ for some $i \in L$ and let us define L to be *inactive* otherwise. Then the probability that $L \cup M$ casts x votes in favor of the policy proposal is

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

I first want to show that

$$P\left[\sum_{i\in L\cup M} v_i = \frac{N_L+N_M}{2} - k|L\ active\right] \geq P\left[\sum_{i\in L\cup M} v_i = \frac{N_L+N_M}{2} + k|L\ active\right]. \tag{4.1}$$

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

$$P\left[\sum_{i\in M} s_i = \frac{N_M-1}{2} - k\right] = P\left[\sum_{i\in M} s_i = \frac{N_M+1}{2} + k\right]$$

for any k, rewrite inequality 4.1 as:

$$\begin{aligned}
&P[\sum_{i\in L} v_i = N_L|L\ active]P[\sum_{i\in M} s_i = \frac{N_M-N_L}{2} - k]\\
&+P[\sum_{i\in L} v_i = 0|L\ active]P[\sum_{i\in M} s_i = \frac{N_M+N_L}{2} - k] \geq\\
&\geq P[\sum_{i\in L} v_i = N_L|L\ active]P[\sum_{i\in M} s_i = \frac{N_M+N_L}{2} - k]\\
&+P[\sum_{i\in L} v_i = 0|L\ active]P[\sum_{i\in M} s_i = \frac{N_M-N_L}{2} - k].
\end{aligned}$$

Regrouping terms:

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

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

Second, I want to show that

$$P\left[\sum_{i\in L\cup M} v_i = \frac{N_L+N_M}{2} - k|L\ inactive\right] \geq P\left[\sum_{i\in L\cup M} v_i = \frac{N_L+N_M}{2} + k|L\ inactive\right]. \quad (4.2)$$

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

$$P\left[\sum_{i\in L\cup M} v_i = \frac{N_L+N_M}{2} + k|L\ ina.\right] - P\left[\sum_{i\in L\cup M} v_i = \frac{N_L+N_M}{2} - k|L\ ina.\right] =$$

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

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

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 (4.2) holds as desired. ■

I now prove lemma 25.

Proof. For any $h \in R$, let A_h

$$= P[\sum_{i\in R_{-h}} s_i = r_R N_R - 1][P\sum_{m\in M\cup L} v_m \in [\frac{N_L+N_M}{2} - \frac{N_R-1}{2}, \frac{N_L+N_M}{2} + \frac{N_R-1}{2} - r_R N_R]]$$

$$-P[\sum_{i\in R_{-h}} s_i \leq (1-r_R)N_R - 1]P[\sum_{m\in M\cup L} v_m = \frac{N_L+N_M}{2} + \frac{N_R-1}{2}]$$

and similarly let $B_h =$

$$P[\sum_{i\in R_{-h}} s_i = (1-r_R)N_R]P[\sum_{m\in M\cup L} v_m \in [\frac{N_L+N_M}{2} - \frac{N_R-1}{2} + r_R N_R, \frac{N_L+N_M}{2} + \frac{N_R-1}{2}]$$
$$-P[\sum_{i\in R_{-h}} s_i \geq r_R N_R]P[\sum_{m\in M\cup L} v_m = \frac{N_L+N_M}{2} - \frac{N_R-1}{2}].$$

Then h prefers to participate in the voting bloc (R, r_R) if and only if $t_h A_h + (1-t_h)B_h > 0$. Suppose $t_l A_l + (1-t_l)B_l > 0$. We want to show that

$$t_h A_h + (1-t_h)B_h - t_l A_l - (1-t_l)B_l \geq 0,$$

which implies $t_h A_h + (1-t_h)B_h > 0$.

Let

$$P_1 = P[\sum_{m\in M\cup L} v_m \in [\frac{N_M+N_L}{2} - \frac{N_R-1}{2}, \frac{N_M+N_L}{2} + \frac{N_R-1}{2} - r_R N_R]],$$

$$P_2 = P[\sum_{m\in M\cup L} v_m \in [\frac{N_M+N_L}{2} - \frac{N_R-1}{2} + r_R N_R, \frac{N_M+N_L}{2} + \frac{N_R-1}{2}]],$$

$$P_3 = P[\sum_{m\in M\cup L} v_m = \frac{N_L+N_M}{2} + \frac{N_R-1}{2}], \text{ and } P_4 = P[\sum_{m\in M\cup L} v_m = \frac{N_M+N_L}{2} - \frac{N_R-1}{2}].$$

Then, $t_h A_h + (1-t_h)B_h$ is equal to:

$$\begin{aligned} &t_h \left[\begin{array}{c} \left(t_l P[\sum_{i\in R_{-lh}} s_i = r_R N_R - 2] + (1-t_l)P[\sum_{i\in R_{-lh}} s_i = r_R N_R - 1] \right) P_1 \\ - \left(t_l P[\sum_{i\in R_{-lh}} s_i \leq (1-r_R)N_R - 2] + (1-t_l)P[\sum_{i\in R_{-lh}} s_i \leq (1-r_R)N_R - 1] \right) P_3 \end{array} \right] \\ &+ (1-t_h) \left[\begin{array}{c} \left(t_l P[\sum_{i\in R_{-lh}} s_i = (1-r_R)N_R - 1] + (1-t_l)P[\sum_{i\in R_{-lh}} s_i = (1-r_R)N_R] \right) P_2 \\ - \left(t_l P[\sum_{i\in R_{-lh}} s_i \geq r_R N_R - 1] + (1-t_l)P[\sum_{i\in R_{-lh}} s_i \geq r_R N_R] \right) P_4 \end{array} \right], \end{aligned} \tag{4.3}$$

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

$$t_l \left[\begin{array}{c} \left(t_h P[\sum_{i \in R_{-lh}} s_i = r_R N_R - 2] + (1-t_h) P[\sum_{i \in R_{-lh}} s_i = r_R N_R - 1] \right) P_1 \\ - \left(t_h P[\sum_{i \in R_{-lh}} s_i \leq (1-r_R) N_R - 2] + (1-t_h) P[\sum_{i \in R_{-lh}} s_i \leq (1-r_R) N_R - 1] \right) P_3 \end{array} \right] \tag{4.4}$$
$$+ (1-t_l) \left[\begin{array}{c} \left(t_h P[\sum_{i \in R_{-lh}} s_i = (1-r_R) N_R - 1] + (1-t_h) P[\sum_{i \in R_{-lh}} s_i = (1-r_R) N_R] \right) P_2 \\ - \left(t_h P[\sum_{i \in R_{-lh}} s_i \geq r_R N_R - 1] + (1-t_h) P[\sum_{i \in R_{-lh}} s_i \geq r_R N_R] \right) P_4 \end{array} \right].$$

Therefore $t_h A_h + (1-t_h)B_h - t_l A_l - (1-t_l)B_l$ is equal to

$$(t_h - t_l) \left(\begin{array}{c} P[\sum_{i \in R_{-lh}} s_i = r_R N_R - 1] P_1 - P[\sum_{i \in R_{-lh}} s_i = (1-r_R) N_R - 1] P_2 \\ + P[\sum_{i \in R_{-lh}} s_i \geq r_R N_R - 1] P_4 - P[\sum_{i \in R_{-lh}} s_i \leq (1-r_R) N_R - 1] P_3 \end{array} \right). \tag{4.5}$$

Since M is symmetric and L leans left, it follows by Claim 41 that $P_1 \geq P_2$ and $P_3 \leq P_4$, and since R_{-h} leans right, by Claim 23 R_{-lh} leans right as well. Then, the expression (4.5) above is weakly positive. ■

4.6.2 Proof of Proposition 26

Proof. By lemma 25, if l prefers to participate in the voting bloc, every member of R does. Therefore, (R, r_R) is Individual-Exit stable if and only if l wants to participate in the bloc. Using the notation from lemma 25, l wants to participate in the bloc if and only if $t_l A_l + (1-t_l)B_l \geq 0$. Suppose $A_l \geq B_l$, then the expression is increasing in t_l and the cutoff that makes the agent indifferent is at $t^{InR}(r_R, r_L, t_{-l}) = \frac{-B_l}{A_l - B_l}$. Hence, it suffices to show that $A_l \geq B_l$.

$$A_l = P[\sum_{i \in R_{-l}} s_i = r_R N_R - 1] P[\sum_{m \in M \cup L} v_m \in [\frac{N_M + N_L}{2} - \frac{N_R - 1}{2}, \frac{N_M + N_L}{2} + \frac{N_R - 1}{2} - r_R N_R]]$$
$$- P[\sum_{i \in R_{-l}} s_i \leq (1-r_R) N_R - 1] P[\sum_{m \in M \cup L} v_m = \frac{N_M + N_L}{2} + \frac{N_R - 1}{2}],$$

$$B_h = P[\sum_{i\in R_{-l}} s_i = (1-r_R)N_R]P[\sum_{m\in M\cup L} v_m \in [\frac{N_M+N_L}{2} - \frac{N_R-1}{2} + r_R N_R, \frac{N_M+N_L}{2} + \frac{N_R-1}{2}]$$
$$-P[\sum_{i\in R_{-l}} s_i \geq r_R N_R]P[\sum_{m\in M\cup L} v_m = \frac{N_M+N_L}{2} - \frac{N_R-1}{2}].$$

Since R_{-l} leans right, for any $r_R \geq \frac{N_R+1}{2N_R}$,

$$P[\sum_{i\in R_{-l}} s_i = r_R N_R - 1] \geq P[\sum_{i\in R_{-l}} s_i = (1-r_R)N_R]$$

and

$$P[\sum_{i\in R_{-l}} s_i \geq r_R N_R] \geq P[\sum_{i\in R_{-l}} s_i \leq (1-r_R)N_R - 1].$$

Since M is symmetric and L leans left,

$$\begin{aligned} P[\sum_{m\in M\cup L} v_m &\in [\frac{N_L+N_M}{2} - \frac{N_R-1}{2}, \frac{N_L+N_M}{2} + \frac{N_R-1}{2} - r_R N_R]] \\ &\geq P[\sum_{m\in M\cup L} v_m \in [\frac{N_L+N_M}{2} - \frac{N_R-1}{2} + r_R N_R, \frac{N_L+N_M}{2} + \frac{N_R-1}{2}] \end{aligned}$$

and

$$P[\sum_{m\in M\cup L} v_m = \frac{N_L+N_M}{2} + \frac{N_R-1}{2}] \leq P[\sum_{m\in M\cup L} v_m = \frac{N_L+N_M}{2} - \frac{N_R-1}{2}].$$

Therefore, $A_l \geq B_l$. ■

4.6.3 Proof of Proposition 28

This proof and the proof of proposition 31 use the result in proposition 32 below. While it is in principle inadvisable to use latter results in proofs that appear earlier in the text, proposition 32 shows that voting blocs generate a gain in utility in a more general model with an endogenous number of voting blocs. To prove the result first for two blocs to use it here and then prove it again in greater generality would be redundant. I also use the notion of a "voting bloc structure" from Definition 13. In short, a voting bloc structure (π, r) is a pair composed of a partition of the assembly π, and a vector r that contains one rule for each voting bloc resulting from the partition π.

Proof. Let (π, r) be a voting bloc structure in which R does not form a voting bloc. Let (π', r') be another voting bloc structure in which R forms a voting bloc with $r_R = \frac{N_R - 1}{N_R}$ and all else remains equal. From proposition 32, $\sum_{i \in R} u_i(\pi', r') \geq \sum_{i \in R} u_i(\pi, r)$. Let (π'', r'') be a third voting bloc structure in which i deviates and leaves the bloc (R, r_R) to become an independent, so the bloc shrinks to $(R \backslash i, r_R)$. Note that the new size of the bloc is $N_R - 1$. Hence the number necessary to command a sufficient majority to roll the minority inside the bloc is

$$r_R(N_R - 1) = N_R - 1 - \frac{N_R - 1}{N_R} > N_R - 2.$$

The new bloc only votes together if the internal majority is of size $N_R - 1$. In other words, r_R is effectively unanimity once i leaves the bloc. Under this rule $R \backslash i$ behaves exactly as if it did not form a bloc and all agents were independent. Thus,

$$\sum_{i \in R} u_i(\pi'', r'') = \sum_{i \in R} u_i(\pi, r) \leq \sum_{i \in R} u_i(\pi', r').$$

Since all agents in R are identical, it follows that for all $i \in R$,

$$u_i(\pi'', r'') \leq u_i(\pi', r').$$

Therefore, no agent wants to leave R and R is Individual-Exit stable. ■

4.6.4 Proof of Proposition 29

Proof. From proposition 28, if $r_R = \frac{N_R - 1}{N_R}$ and R is homogeneous, then (R, r_R) is Individual-Exit stable. Hence it suffices to show that there exist a homogeneous type profile for R such that with simple majority the bloc is not stable. Let $P[\sum_{i \in M \cup L} v_i = \frac{N+1}{2} - N_R] = \lambda$. By the assumption on sizes and types of M and L, $\lambda > 0$. Let the common type of agents in R be $1 - \varepsilon$. Let E be the event that $i \in R$ rejects the proposal, a majority of R favors the proposal, and $\sum_{i \in M \cup L} v_i = \frac{N+1}{2} - N_R$. In this event, i is better off if she is not part of the bloc. Note that

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} P[E] = \lambda.$$

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

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} P[\sum_{j \in R} s_j = \frac{N_R - 1}{2}] = 0.$$

Therefore, for a sufficiently low ε the probability that i is better off outside the bloc outweighs the probability that i is better off inside the bloc, and i prefers to leave the bloc. ■

4.6.5 Proof of Proposition 30

Proof. Let x_h be the highest coordinate of the vector x and let $\varepsilon = \alpha x_h$ be the highest type in L. L is stable if $\varepsilon A_h + (1 - \varepsilon) B_h > 0$.

$$\begin{aligned} & A_h \\ = {} & P\left[\sum_{i \in L_{-h}} s_i = r_L N_L - 1\right] P\left[\sum_{m \in M \cup L} v_m \in \left[\frac{M + N_R}{2} - \frac{N_L - 1}{2}, \frac{M + N_R}{2} + \frac{N_L - 1}{2} - r_L N_L\right]\right] \\ & - P\left[\sum_{i \in L_{-h}} s_i \leq (1 - r_L) N_L - 1\right] P\left[\sum_{m \in M \cup R} v_m = \frac{M + N_R}{2} + \frac{N_L - 1}{2}\right]. \end{aligned}$$

$$B_h$$

$$= P\left[\sum_{i\in L_{-h}} s_i = (1-r_L)N_L\right] P\left[\sum_{m\in M\cup R} v_m \in \left[\frac{M+N_R}{2} - \frac{N_L-1}{2} + r_L N_L, \frac{M+N_R}{2} + \frac{N_L-1}{2}\right]\right]$$
$$-P\left[\sum_{i\in L_{-h}} s_i \geq r_L N_L\right] P\left[\sum_{m\in M\cup R} v_m = \frac{M+N_R}{2} - \frac{N_L-1}{2}\right].$$

Let

$$P_5 = P\left[\sum_{m\in M\cup R} v_m \in \left[\frac{M+N_R}{2} - \frac{N_L-1}{2}, \frac{M+N_R}{2} + \frac{N_L-1}{2} - r_L N_L\right]\right],$$

$$P_7 = P\left[\sum_{m\in M\cup R} v_m \in \left[\frac{M+N_R}{2} - \frac{N_L-1}{2} + r_L N_L, \frac{M+N_R}{2} + \frac{N_L-1}{2}\right]\right],$$

$$P_6 = P\left[\sum_{m\in M\cup R} v_m = \frac{M+N_R}{2} + \frac{N_L-1}{2}\right] \text{ and } \gamma = \frac{(N_L-1)!}{(r_L N_L-1)!(N_L-r_L N_L)!}.$$

Then,

$$\varepsilon A_h + (1-\varepsilon)B_h < \varepsilon\gamma\varepsilon^{r_L N_L-1}P_5 - \varepsilon(1-\varepsilon)^{N_L-1}P_6 + (1-\varepsilon)\gamma\varepsilon^{(1-r_L)N_L}P_7.$$

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

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

Hence, if ε is low enough, $\varepsilon A_h + (1-\varepsilon)B_h < 0$ and the voting bloc is not Individual-Exit stable. ■

4.6.6 Proof of Proposition 31

Proof. Let the vector of types be such that $t_i = t_j$ for all $i, j \in L$. Then, by proposition 32, every $i \in L$ weakly benefits from the formation of the voting bloc (L, r_L) and hence it is a weakly undominated strategy for every $i \in L$ to commit to participate in the bloc. By proposition 30, if the common type of L is low enough, (L, r_L) is not Individual-Exit stable. ■

4.6.7 Proof of Proposition 32

Proposition 32 and its proof follow very closely proposition 12 in chapter 3, extending the result from one to several voting blocs.

Proof. Let (π, r) be the initial voting bloc structure and (π', r') the new voting bloc structure in which $r' = \{r_j\}_{j=0}^{J} \cup r'_{J+1}$ and $\pi' = \{C'_j\}_{j=0}^{J+1}$ is a finer partition of π such that $C'_{J+1} \cup C'_0 = C_0$ and $C'_j = C_j$ for all $j = \{1, ..., J\}$. For notational simplicity, let C'_{J+1}be just C' and r'_{J+1}simply r'_C.

Given s, suppose $\sum\limits_{i \in C'} s_i \leq (1 - r'_C)N_{C'}$. Then $\sum\limits_{i \in C'} v_i(\pi', r', s) \leq \sum\limits_{i \in C} v_i(\pi, r, s)$ and since the votes of other agents are unaffected by the formation or not of a bloc (C', r'_C), it follows that $\sum\limits_{i \in \mathcal{N}} v_i(\pi', r', s) \leq \sum\limits_{i \in \mathcal{N}} v_i(\pi, r, s)$. Hence, either the outcome is the same under (π, r) and (π', r'), or if the outcome changes, it must be that the policy proposal passes under (π, r) but fails under (π', r') and then every agent who is against the proposal benefits from the formation of the bloc (C', r'_C) and every agent who likes the policy proposal is hurt. If the outcome changes, the aggregate gain in utility for the coalition C' is equal to $N_{C'} - 2\sum\limits_{i \in C'} s_i \geq (2r'_C - 1)N_{C'} \geq 1$.

Suppose instead that $(1 - r'_C)N_{C'} < \sum\limits_{i \in C'} s_i < r'_C N_{C'}$. Then the formation of the voting bloc (C', r'_C) does not affect the voting behavior, the policy outcome or the utility of any agent.

Finally, suppose that $\sum\limits_{i \in C'} s_i \geq r'_C N_{C'}$. Then, by a symmetric logic to the one in the first case, the outcome can only change from rejecting to accepting the policy proposal, which benefits a majority of members of the new bloc.

Hence, either the bloc has no effect, or if it has an effect, it generates a strictly positive surplus of utility for its members.

Simple majority maximizes this surplus because with simple majority the bloc always rolls its internal minorities, maximizing the number of preference profiles s for which it alters the outcome in favor of the majority of the voting bloc. ■

4.6.8 Proof of Proposition 33

Proof. If $\exists\, i \in \mathcal{N} \backslash C'_{J+1}$ s.t. $v_i(\pi', r', s) = 0$, then it is irrelevant whether the coalition C'_{J+1} forms a voting bloc as in the structure (π', r'), or it does not as in (π, r). In either case the outcome is a rejection of the policy proposal. On the other hand, if $v_i(\pi', r', s) = 1$ for every $i \in \mathcal{N} \backslash C'_{J+1}$, then if C'_{J+1} does not form a voting bloc and a single member of the coalition votes against the proposal, the proposal fails; whereas, if C'_{J+1} forms a voting bloc the proposal only fails if strictly more than $(1 - r'_{J+1})N_{J+1}$ members of C'_{J+1} are against it. By assumption, $r'_{J+1} < 1$ and $r'_{J+1}N_{J+1}$ is an integer, so $r'_{J+1} \leq \frac{N_{J+1}-1}{N_{J+1}}$ and $(1 - r'_{J+1})N_{J+1} \geq 1$. Since Ω has full support, there is some positive probability that exactly one member of the bloc opposes the policy proposal and then formation of the bloc (C'_{J+1}, r'_{J+1}) alters the outcome, from a rejection of a proposal favored by every $i \in C'_0$, to an acceptance. Therefore, $i \in C'_0$ benefits from the formation of the bloc C'_{J+1}. Social welfare increases because the only case in which the formation of a bloc by C'_{J+1} affects the outcome is if a sufficient majority of every bloc including (C'_{J+1}, r'_{J+1}) and every singleton, but not every member of C'_{J+1} favor the policy proposal. In this case the policy proposal fails if C'_{J+1} does not form a bloc, and passes if it forms a bloc; a majority of the assembly (indeed, a sum of majorities in every bloc) prefers the outcome that occurs if C'_{J+1} forms a bloc. $\blacksquare$

4.6.9 Proof of Proposition 34

Proof. First I show that the voting bloc structures described in the proposition are relevant, then that they are Individual-Exit stable, and finally that at least one of them exists.

Suppose $\sum\limits_{i \in C} s_i = \sum\limits_{i \in \mathcal{N}} s_i = r_C N_C < r_{\mathcal{N}} N$ so the policy proposal fails if $v_i = s_i$ for all $i \in \mathcal{N}$. However, if the voting bloc (C, r_C) forms, the proposal wins the internal voting of the bloc, $\sum\limits_{i \in C} v_i = \sum\limits_{i \in \mathcal{N}} v_i = N_C \geq r_{\mathcal{N}} N$ and the proposal passes in the division of the assembly. Since $\Omega(s)$ has full support, $\sum\limits_{i \in C} s_i = \sum\limits_{i \in \mathcal{N}} s_i = r_C N_C$ occurs with positive probability and the bloc is relevant.

Since by assumption $N_C - 1 \geq r_{\mathcal{N}} N$, the bloc remains a dictator after losing one member.

Suppose $i \in C$ and $v_i = s_i$, agent i is at least equally well off staying in the bloc since i is already voting her preference and by leaving she can never increase the number of other agents who vote for her preference in the division of the assembly. Suppose $i \in C$ and $v_i \neq s_i$. Then it must be that i lost in the internal vote of the bloc, and $v_j \neq s_i$ for all $j \in C$. If i leaves the bloc, it would still be that a sufficient majority of members of C oppose $i's$ preference, and $v_j \neq s_i$ for all $j \in C \backslash i$. Since the bloc without i remains a dictator, i still loses in the division of the assembly after her defection from the bloc. Therefore, agent i can never be better off leaving the bloc and the bloc is Individual-Exit stable.

Finally, I want to show that for any $r_{\mathcal{N}} \leq \frac{N-1}{N}$ and any $N \geq 7$ there exists an r_C and N_C such that $r_C > \frac{1}{2}$, $r_C N_C$ is an integer, $N_C \geq r_{\mathcal{N}} N + 1$ and $r_C N_C < r_{\mathcal{N}} N$ so that the second statement in the proposition applies. This is straightforward: If $r_{\mathcal{N}} = \frac{N+1}{2N}$, let $N_C = N - 2$ and $r_C = \frac{N-1}{2(N-2)}$, and if $r_{\mathcal{N}} > \frac{N+1}{2N}$, let $r_C = \frac{N+1}{2N}$ and $N_C = N$. ∎

4.6.10 Proof of Proposition 35

Proof. First consider the with $r_{\mathcal{N}} \in (\frac{N+1}{2N}, \frac{N-1}{N}]$. Then, by proposition 34, any voting bloc structure with a unique voting bloc (C, r_C) such that $C = \mathcal{N}$ and $r_C < r_{\mathcal{N}}$ is relevant and Individual-Exit stable. Since there are no agents outside the bloc, there is no possible deviation by entering the bloc and the voting bloc structure is also Individually stable.

Suppose instead that $r_{\mathcal{N}} = \frac{N+1}{2N}$. Let (π, r) be any voting bloc structure with a unique voting bloc (C, r_C) such that $N_C = N - 2$ and $r_C = \frac{N_C+1}{2N_C}$ so that $r_C N_C = \frac{N_C+1}{2} = \frac{N-1}{2}$. By proposition 34, the voting bloc structure is relevant and Individual-Exit stable, so the only deviations that need to be ruled out are those by a non-member who enters the bloc. Suppose a non-member l deviates and enters the bloc, so that the new bloc is now $(C \cup l, r_C)$. The deviation affects the outcome in the division of the assembly only if $\sum_{i \in C \cup l} s_i = \frac{N-1}{2}$. In this case, the result in the new bloc is a tie. Without l, the result was an internal majority of 1 against the preference of l and the whole bloc casting all its votes against the preference of l in the division of the assembly. If by entering the bloc and bringing a tie inside the bloc l reverts the outcome in the division of the assembly, then a majority of members of C are hurt by the inclusion of l. Thus, there is a net loss of utility for the members of C. It must then be that in expectation at least one of them is ex ante worse off by the entry of agent l, so member l cannot deviate by entering. Consider the incentives of any $i \in C$ to leave the bloc. For any s such that $v_i(\pi, r, s) = s_i$ member i is at least equally well off staying in the bloc. Therefore, the bloc is Individually stable. ■

4.6.11 Proof of Proposition 37

Proof. The grand coalition $C = \mathcal{N}$ with $r_C \in [r_{\mathcal{N}}, 1]$ is irrelevant, but Nash stable. For any s such that agent $v_i(\mathcal{N}, r_C, s) = s_i$, agent i is at least equally well off remaining in the bloc. For any s such that $s_i = 0$ but $v_i(\mathcal{N}, r_C, s) = 1$ it must be that $\sum\limits_{j \in \mathcal{N} \setminus i} s_j \geq r_C N \geq r_C(N-1)$ so if i leaves the bloc, all $N-1$ members vote in favor of the proposal and the proposal passes, so i is not better off. For any s such that $s_i = 1$ but $v_i(\mathcal{N}, r_C, s) = 0$ it must be that $\sum\limits_{j \in \mathcal{N} \setminus i} s_j \leq (1-r_C)N - 1 \leq (1-r_C)(N-1)$ so if i leaves the whole bloc votes against the proposal, the proposal fails and i is not better off. Overall, an agent can never change the outcome towards her preference by leaving the grand coalition, so $(\mathcal{N}, r_C)$ with $r_C \in [r_{\mathcal{N}}, 1]$ is Nash stable. ■

4.6.12 Proof of Proposition 38

Proof. By contradiction. Suppose $\exists (C, r_C)$ such that r_C is simple majority and $\frac{N+1}{2} \leq N_C$. If $N_C = N$, then the bloc is not relevant -a contradiction. Suppose $N_C < N$. For any s such that $\sum\limits_{h \in C} s_h \neq \frac{N_C}{2}$, it follows that $\sum\limits_{h \in C} v_h \in \{0, N_C\}$ and the policy outcome in the division of the assembly coincides with the vote of the bloc; since the policy outcome is independent of the votes outside the voting bloc, any $i \notin C$ is at least equally well off entering the voting bloc. For any s such that $\sum\limits_{h \in C} s_h = \frac{N_C}{2}$, any $i \notin C$ who joins the bloc causes $\sum\limits_{h \in C \cup i} v_h = s_i N_C$ and i wins in the division of the assembly with all the votes of the bloc; if i was winning outside of the bloc, i is indifferent between winning outside the bloc or being pivotal to win inside the bloc, and if i was losing, i is strictly better off entering the bloc. There is no preference profile s for which an agent i is better off staying out of the bloc.

If the bloc is odd sized, ties cannot occur. To find a case in which i is strictly better off entering a bloc of odd size, let s be such that $\sum_{h\in C} s_h = \sum_{h\in\mathcal{N}} s_h = \frac{N_C+1}{2}$. Then $\sum_{h\in C} v_h = \sum_{h\in\mathcal{N}} v_h = N_C \geq \frac{N+1}{2}$ and the proposal passes. If one of the non-members -who oppose the proposal- joins the bloc, then the expanded bloc is tied, $\sum_{i\in\mathcal{N}} v_i = \sum_{h\in C\cup i} v_h = \sum_{h\in C} s_h < \frac{N+1}{2}$ and the proposal does not pass in the assembly. Therefore, regardless of whether ties can occur or not in the bloc, for any non-member i there exist preference profiles 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. Then, if $C \neq \mathcal{N}$, the voting bloc structure is not Nash stable -a contradiction.

Suppose the voting bloc structure (π, r) is such that $C_0 \neq \emptyset$, and $\exists(C, r_C)$ relevant such that $N_C \geq \frac{N+1}{2}$. Let (π', r) be a new voting bloc structure such that $C' = C\cup i$ and $C_0' = C_0\backslash i$ and all else is unchanged. Let s be a preference profile such that $v_i(\pi', r, s) = s_i$. Then $u_i(\pi', r, s) \geq u_i(\pi, r, s)$ since i joining the bloc can never reduce the number of votes cast by other bloc members for the option preferred by i. Suppose instead that s is such that $v_i(\pi', r, s) \neq s_i$. Since the bloc is a dictator in the assembly, then $u_i(\pi', r, s) = 0$. But note that the bloc would also vote against i if i remained out of the bloc. Since the bloc without i is also a dictator, $u_i(\pi, r, s) = 0$. So the agent is in this case indifferent about joining the bloc. In either case, an agent is never worse off joining the bloc. Let s be such that $\sum_{h\in C} s_h = (1 - r_C)N_C$ and $s_k = 1$ for all $k \notin C$. Then $\sum_{h\in C} v_h = 0$ and the proposal fails in the division of the assembly. Since the voting bloc is relevant, $\sum_{h\in C} s_h + \sum_{k\notin C} s_k \geq \frac{N+1}{2}$ and the proposal would pass if the members of the voting bloc voted sincerely in the assembly. Suppose $i \notin C$ enters the bloc, so that $C' = C \cup i$. Then $\sum_{h\in C'} s_h > (1 - r_C)N_{C'}$ and $\sum_{h\in C} v_h = \sum_{h\in C} s_h$ so that the policy proposal passes in the division of the assembly and i is better off -a contradiction. ■

4.6.13 Estimates by Eigen-D, SVD and OCM-2D

The top table contains the first and second eigenvectors obtained by the Eigen Decomposition of the double-centered matrix of squared distances of the justices, the estimates of the location of the justices in the first and second dimension by Single Value Decomposition and the estimates of the location in the first and second dimensions obtained by the Optimal Classification method with two dimensions. Note that the first dimension with SVD is an "agreement dimension" where all justices take a similar position, and it is only the second dimension that is the relevant and meaningful one, comparable to the first dimension in the other methods. The bottom table provides the nine single values of the Eigen Decomposition, and the weights of the nine dimensions from the SVD (to obtain the single value of each dimension, multiply by 11.065).

Table 4.5: Estimates by Eigen-D, SVD and OCM.

Single values	1.Ste	2.Gin	Sou	Bre	5.O'C	6.Ken	7.Reh	Sca	Tho
1st-Eigenv.	−0.42	−0.30	−0.250	−0.253	0.161	0.21	0.348	0.455	0.459
2nd-Eigenv.	0.339	0.026	0.198	−0.467	−0.49	−0.15	−0.35	0.277	0.406
SVD 1stD	−0.24	−0.31	−0.331	−0.326	−0.40	−0.40	−0.36	−0.30	−0.30
SVD 2ndD	−0.42	−0.38	−0.351	−0.335	0.089	0.154	0.294	0.398	0.402
OCM 1stD	−0.74	−0.58	−0.506	−0.498	0.169	0.274	0.489	0.704	0.661
OCM 2ndD	0.569	0.087	0.387	−0.730	−0.71	−0.18	−0.52	0.482	0.661

Table 4.6: Relative weights of the nine dimensions

Dimension i	$1st$	$2nd$	$3rd$	$4th$	$5th$	$6th$	$7th$	$8th$	$9th$
Eigenvalue α_i	−0.817	0.082	0.044	0.037	0.022	−0.021	0.015	−0.008	0.006
Weight Dim SVD	0.385	0.179	0.079	0.072	0.067	0.062	0.056	0.052	0.047

4.6.14 Table following Result 40

The first column in each row contains the voting bloc structure as a list of the blocs that form; the numbers inside each bloc correspond to the justices in the order given in the top row. The other cells detail the payoff to each justice. The first two voting bloc structures are Split and Nash Stable voting bloc structures. The others are not.

Table 4.7: The effect of different voting bloc structures

Bloc structure	1.Ste	2.Gin	3.Sou	4.Bre	5.O'Co	6.Ken	7.Reh	8.Sca	9.Tho
$\{124\}, \{789\}$	-1	-5	-1	-1	9	-3	-7	-1	-7
$\{134\}, \{789\}$	-1	1	-7	-3	11	-3	-9	-1	-5
$\{345\}$	75	77	71	85	-67	-79	-77	-85	-89
$\{456\}$	-4	-4	-10	6	14	2	-2	-8	-16
$\{567\}$	-50	-48	-52	-44	-12	16	48	46	44
$\{4567\}$	-8	-4	-10	-2	6	6	10	0	-8
$\{34567\}$	-15	-15	-11	7	17	9	11	-15	-23

Bibliography

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

[2] Aldrich, John H. and David W. Rohde. 2001. "The logic of conditional party government: Revisiting the electoral connection," in *Congress Reconsidered, 7th Ed.* CQ Press.

[3] Alvarez, Michael and Jonathan Nagler. 2000. "A new approach for modeling strategic voting in multiparty elections," *British Journal of Political Science*, *30*: 57—75.

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

[5] Badger, Wade W. 1972. "Political individualism, positional preferences, and optimal decision rules," in *Probability Models of Collective Decision Making*. Merril Publishing.

[6] Barberá, Salvador and Matthew Jackson. 2004. "Choosing how to choose: Self-stable majority rules," *The Quarterly Journal of Economics*, *119* (3): 971—1009.

[7] Baron, David. 1993. "Government formation and endogenous parties," *American Political Science Review*, *85* :137—164.

[8] Bernheim, B. Douglas, Bezalel Peleg and Michael D. Whinston. 1987. "Coalition-proof Nash equilibria: I. Concepts," *Journal of Economic Theory*, *42* (1): 1—12.

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

[10] Bloch, Francis. 2003. "Non-cooperative models of coalition formation in games with spillovers," in *The Endogenous Formation of Economic Coalitions*. Edward Elgar Publishing.

[11] Brazill, Timothy J. and Bernard Grofman. 2002. "Factor analysis versus multi-dimensional scaling: Binary choice roll-call voting and the US Supreme Court," *Social Networks*, *24*: 201—229.

[12] Buchanan, James M. and Gordon Tullock. 1962. *The Calculus of Consent*. Ann Arbor: University of Michigan Press.

[13] Cadigan, John. 2005. "The citizen candidate model: An experimental analysis." *Public Choice*, *123*: 197—216.

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

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

[16] Curtis, Richard B. 1972. "Decision rules and collective values in constitutional choice," in *Probability Models of Collective Decision Making*. Merril Publishing.

[17] Darroch, John N. 1964. "On the distribution of the number of successes in independent trials," *The Annals of Mathematical Statistics*, *35* (3): 1317—1321.

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

[19] Drèze, Jacque and Joseph Greenberg. 1980. "Hedonic coalitions: Optimality and stability," *Econometrica*, *48* (4): 987—1003.

[20] Eckart, Carl and Gale Young. 1936. "The approximation of one matrix by another of lower rank," *Psychometrika*, *1*: 211—218.

[21] Eguia, Jon X. 2006. "United we vote." FEEM Working Paper 9.2006.

[22] Eguia, Jon X. Forthcoming. "Citizen candidates under uncertainty," *Social Choice and Welfare*.

[23] Eguia, Jon X. Forthcoming. "United we vote," *Journal of Public Economic Theory*.

[24] Epstein, Lee and Jack D. Knight. 1998. *The Choices Justices Make*. Washington DC: CQ Press.

[25] Epstein, Lee, Andrew D. Martin and Kevin M. Quinn. 2005. "The Rehnquist court?" *Law and Courts*, *15*: 18—23.

[26] Feix, Marc R., Dominique Lepelley, Vincent R. Merlin and Jean-Louis Rouet. 2007. "On the voting power of an alliance and the subsequent power of its members," *Social Choice and Welfare*, *28*: 181—207.

[27] Felsenthal, Dan S. and Moshe Machover. 2002. "Annexations and alliances: When are blocs advantageous a priori," *Social Choice and Welfare*, *19*: 295—312.

[28] 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).

[29] Hart, Sergio and Mordechai Kurz. 1984. "Stable coalition structures," in *Coalitions and Collective Action*. Physica Verlag.

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

[31] Hosli, Madeleine O. 2000. "Coalition formation and voting in the Council of the European Union," *Homo Oeconomicus*, *17*: 31—55.

[32] Hosli, Madeleine O. 2002. "Preferences and power in the European Union," *Homo Oeconomicus*, *19* (3): 311—326.

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

[34] Kaminski, Marek M. 2001. "Coalitional stability of multi-party systems: Evidence from Poland," *American Journal of Political Science*, *45*: 294—312.

[35] Kaminski, Marek M. 2006. "A general equilibrium model of multi-party competition," *Social Choice and Welfare*, *26*: 333—361.

[36] Laruelle, Annick and Federico Valenciano. 2005. "Assessing success and decisiveness in voting situations," *Social Choice and Welfare*, *24*: 171—197.

[37] Lee, David S., Enrico Moretti and Matthew Butler. 2004. "Do voters affect or elect policies? Evidence from the US House," *Quarterly Journal of Economics*, *119*: 807—859.

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

[39] Maggi, Giovanni and Massimo Morelli. 2006. "Self enforcing voting in international organizations," *American Economic Review*, *96* (4): 1137—1158.

[40] Martin, Andrew D. and Kevin M. Quinn. 2002. "Dynamic ideal point estimation via Markov chain Monte Carlo for the U.S. Supreme Court, 1953—1959," *Political Analysis*, *10* (2): 134—153.

[41] McKelvey, Richard and Thomas Palfrey. 1995. "Quantal response equilibria in normal form games," *Games and Economic Behavior*, *10*: 6—38.

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

[43] Myerson, Roger. 1993. "Effectiveness of electoral systems for reducing government corruption: A game-theoretic analysis," *Games and Economic Behavior*, *5*: 118—132.

[44] Osborne, Martin and Al Slivinski. 1996. "A model of political competition with citizen candidates," *Quarterly Journal of Economics*, *111*: 65—96.

[45] Poole, Keith. 2005. *Spatial Models of Parliamentary Voting*. Cambridge University Press.

[46] Poole, Keith and Howard Rosenthal. 1985. "A spatial model for roll call analysis," *American Journal of Political Science*, *29*: 331—347.

[47] Rae, Douglas W. 1969. "Decision rules and individual values in constitutional choice," *American Political Science Review*, *63* (1): 40—56.

[48] Ray, Debraj and Rajiv Vohra. 1997. "Equilibrium binding agreements," *The Journal of Economic Theory*, *73*: 30—78.

[49] Riker, William H. 1962. *The Theory of Political Coalitions*. Yale University Press.

[50] Riviere, Anouk. 2000. "Citizen candidacy, party formation and Duverger's law." Royal Holloway dp00/1 University of London.

[51] Roemer, John. 2003. "Indeterminacy of citizen-candidate equilibrium." Cowles Foundation Discussion Paper 1410.

[52] Rohde, David W. 1972. "Policy goals and opinion coalitions in the Supreme Court," *The Midwest Journal of Political Science*, *16* (2): 208—224.

[53] Rohde, David W. 1991. *Parties and Leaders in the Postreform House*. Chicago: University of Chicago Press.

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

[55] Segal, Jeffrey A. and Harold J. Spaeth. 2002. *The Supreme Court and the Attitudinal Model Revisited*. Cambridge University Press.

[56] Sirovich, Lawrence. 2003. "A pattern analysis of the second Rehnquist U.S. Supreme Court," *Proceedings of the National Academy of Sciences*, *100* (13): 7432—7437.

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

[58] Spaeth, Harold. 2006. "United States Supreme Court Judicial Database 1953—2005." Computer File available at www.as.uky.edu/polisci/ulmerproject, apr 2006.

[59] Yi, Sang-Seung. 2003. "Endogenous formation of economic coalitions: A survey of the partition function approach," in *The Endogenous Formation of Economic Coalitions*. Edward Elgar Publishing.