Voting Blocs, Coalitions and Parties∗

Jon X Eguia†
New York University

May 17, 2007

Abstract

In this paper 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 thank Francis Bloch, Anna Bogomolnaia, Matias Iaryczower, Matt Jackson, Andrea Mattozzi and Tom Palfrey for their comments and suggestions.
†Email: jon@hss.caltech.edu. Mail: 19 West 4th St, 2nd Floor, New York, NY 10012.
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 paper, 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.

First part 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 paper 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.

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

In the coalition formation literature, the seminal work of Buchanan and Tullock [8] analyzes the costs and benefits of forming a coalition and praises the virtues of unanimity as internal voting rule. Hart and Kurz [18] study the endogenous formation of economic coalitions under the restriction that the overall partition of the society into coalitions is efficient. Carraro [9] 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 the coalition they belong to, not by the actions of other agents outside their coalition. In contrast, the partition function approach first used by Thrall and Lucas [31] 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 [6] and Yi [32] 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. More recently, Bloch and Gomes [7] propose a general model to cover a variety of applications with either positive or negative externalities. However, there is no literature yet on the more general case in which a coalition generates both positive and negative externalities to non-members. A forthcoming paper by Hyndman and Ray [21] makes the first contribution to this future literature in a restricted model with only three agents. The formation of a voting bloc or a political party of any size 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 of finitely many agents.
In previous formal theories of party formation, Snyder and Ting [30] describe parties as informative labels that help voters to decide how to vote, Levy [23] stresses that parties act as commitment devices to offer a policy platform that no individual candidate could credibly stand for, Morelli [25] 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 [4] and Jackson and Moselle [22] 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 [10] 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 [3] 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 [24] 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 [27] showed the advantages of forming minimal winning coalitions, Diermeier and Merlo [13] 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 [29] 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 [17] 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.

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 1** 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 with simple majority, so if any 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 1 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 8) 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 14).

In Example 1, 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.\footnote{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.} 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 1 and in much of the voting power literature\footnote{Within this literature, see Felsenthal and Machover [16] for a study of voting blocs, and Hosli [19] and [20] 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.} is an extreme case of uncertainty, when agents not only do not know their exact preferences about future policy proposals, but they cannot even take a guess. In my model, I assume that there is some uncertainty about preferences, but that ex ante it is possible to differentiate agents according to their prior over preferences. For instance, a Republican legislator in the US Senate is ex-ante more likely to favor a future proposal on tax cuts than a Democratic senator, even if the ex-post preferences are not perfectly known.

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 2 Let \( N \) be an assembly with seven agents who must make a binary choice decision -pass or reject some policy proposal- by simple majority. Let each agent \( i \) favor the proposal with an independent probability \( t_i \). Suppose \( t_1 = t_2 = t_3 = \frac{1}{4} \) and \( t_4 = t_5 = t_6 = t_7 = \frac{3}{4} \). Table 1 below shows the probability that the policy proposal 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>
<thead>
<tr>
<th>Voting Blocs</th>
<th>Probability of Policy Proposal Passing</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Blocs</td>
<td>0.25</td>
</tr>
<tr>
<td>Agents 5, 6, 7 form a bloc</td>
<td>0.75</td>
</tr>
<tr>
<td>Agents 1, 2, 3 form a bloc</td>
<td>0.75</td>
</tr>
<tr>
<td>Both blocs form</td>
<td>0.75</td>
</tr>
<tr>
<td>Agents 1, 2, 3 form a bloc and 4, 5, 6, 7 form another bloc</td>
<td>0.75</td>
</tr>
</tbody>
</table>

The numbers on the table come from simple binomial calculations. 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. 
Table 1: Probabilities that the proposal passes and the agents like the outcome

<table>
<thead>
<tr>
<th>Bloc</th>
<th>Pass 7 favors</th>
<th>Pass 7 satisfied</th>
<th>4 satisfied</th>
<th>1 satisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>59.4%</td>
<td>68.8%</td>
<td>68.8%</td>
<td>57.3%</td>
</tr>
<tr>
<td>{5,6,7}</td>
<td>75.7%</td>
<td>83.9%</td>
<td>75.3%</td>
<td>41.9%</td>
</tr>
<tr>
<td>{1,2,3}</td>
<td>42.3%</td>
<td>51.2%</td>
<td>59.5%</td>
<td>63.4%</td>
</tr>
<tr>
<td>{1,2,3},{5,6,7}</td>
<td>68.4%</td>
<td>74.7%</td>
<td>68.6%</td>
<td>45.4%</td>
</tr>
<tr>
<td>{1,2,3},{4,5,6,7}</td>
<td>77.1%</td>
<td>86.6%</td>
<td>77.8%</td>
<td>39.4%</td>
</tr>
</tbody>
</table>

Note that agent 4 in Example 2, 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 paper generalizes the intuitions provided in the example.

**Example 3** 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. The Coalition Caucus opposes any bill 10 – 30 in the internal vote. If all the members of the voting bloc then vote together against the bill in the division of the Upper House, the bill is rejected with only 50 votes, all coming from liberal legislators.

The formation of a voting bloc by a minority in Example 3 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 3 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 2 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.

Second, suppose that the Coalition Caucus changes its rules and adopts the following supermajority internal voting rule: Members must vote together as a bloc only if three quarters of the them share the same preference, otherwise, each member may vote her own preference. In a bloc of either 39 or 40 members this rule requires that at least 30 members share the same preference before the minority is forced to vote with the majority. If all moderates stay in the Coalition, the opponents of the bill always reach the threshold and the bloc functions just as if it was using simple majority, defeating every bill. 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 5: 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 harms 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 blocs. The rest of the paper 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.

3 The Model

Let $N = \{1, 2, ..., N\}$ 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 $i \in N$ 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 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

Agents face uncertainty at the initial stage. They do not know the profile of preferences 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.

Definition 1 Types are independent if $\Omega(s) = \prod_{i \in N} \left[ t_i s_i + (1 - t_i)(1 - s_i) \right]$ 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 $i \in N$ 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
$N = L \cup R \cup M$. 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 $\sum_{i \in C} s_i \geq r_CN_C$, 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 $(1-r_C)N_C < \sum_{i \in C} s_i < r_CN_C$, 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 $r_LN_L$ and $r_RN_R$ 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 $r_CN_C$ 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 consid-
eration 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 all 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 at the constitutional stage. 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.

**Definition 2.** 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 \((C \setminus i, r_C)\) 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.

\(^3\)To be formally precise, I would need to define a new variable \( \hat{s}_i \) to denote the preference expressed by \( i \)
in the internal meeting of coalition \( C \), and let \( \sum_{i \in C} \hat{s}_i \) determine the outcome of the internal meeting, but since sincere
voting is weakly dominant, \( \hat{s}_i = s_i \) for all \( i \in C \), all \( C \subseteq \mathcal{N} \) and all \( s \in S \).
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 in the next section, providing a more formal extension of Definition 2 to encompass the incentives to deviate by all agents in a setting with 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 $\Omega$ over preference profiles, for any $C \subseteq N$, let $g_C(x)$ be the probability that $\sum_{i \in C} s_i = x$. For any $i, h \in C$, let $g_C^{-i}(x)$ be the probability that $\sum_{k \in C \setminus i} s_k = x$ and let $g_C^{-ih}(x)$ be the probability that $\sum_{k \in C \setminus \{i, h\}} s_k = x$.

Definition 3 A set of agents $C$ of odd size $N_C$ leans right if for any non-negative $k$

$$g_C\left(\frac{N_C - 1}{2} - k\right) \leq g_C\left(\frac{N_C + 1}{2} + k\right).$$

$C$ 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:

$$g_C\left(\frac{N_C}{2} - k\right) \leq g_C\left(\frac{N_C}{2} + k\right) \text{ for any positive } k.$$

Definition 3 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.

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 1 Let $N = L \cup M \cup R$. 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 1 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 1 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:

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

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 1 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 2** Let \(N = L \cup M \cup R\). 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 \(t^{InL}(r_L, r_R, t_{-l})\). Taking both results together, a corollary follows:
Corollary 3 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

$$t_h \leq t^{lnL}(r_R,r_L,t_{-h}) \text{ and } t_l \geq t^{lnR}(r_L,r_R,t_{-l}).$$

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 $t^{lnR}(r_R,r_L,t_{-l})$ 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 8 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 4 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 8. 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 5** Let \( \mathcal{N} = L \cup M \cup R \). Suppose that types are independent with \( t_i \in (0,1) \) for all \( i \in \mathcal{N} \setminus R \). Suppose that \( L \) leans left and forms a voting bloc \((L,r_L)\), \( M \) is symmetric, \( N_L \leq \frac{N_R - 1}{2} \) and \( 3 < N_R \leq \frac{N_R + 1}{2} \). 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 1 illustrates this result. To be able to plot \( t^\text{InR}(r_R,r_L,t_{-l}) \) as a function of a single parameter, I let \( N_L = 11, N_R = 9, N_M = 31, t_i = 0.3 \) for all \( i \in L \), \( r_L = \frac{6}{11} \) (simple majority), \( t_m = 0.5 \) for all \( m \in M \) and \( t_h = t_R \) for all \( h \in R \). Given these values, I plot \( t^\text{InR}(r_R,r_L,t_{-l}) \) 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 \( r - \text{majority} \) 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 a supermajority.

This result contrasts with the findings of Maggi and Morelli [24] 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

---

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.
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 [28] 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 [10], page 155:

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

Proposition 5 shows that this finding is not an idiosyncrasy of the Democratic and Republican parties in the US Congress, but rather, a general principle 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 4.

**Proposition 6** Let $N = L \cup M \cup R$. 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 $\alpha$ is low enough, $(L, r_L)$ with $r_L \leq \frac{N_L}{N_L - 2}$ is not Individual-Exit stable.

A corollary to proposition 6 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 4 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 4** Let $N = L \cup M \cup R$. 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 $r_C = \frac{N_C + 1}{2 N_C}$ for $C = L, R$, for a symmetric assembly where $N_L = N_R$ (column two) and an asymmetric assembly where $N_L = 2 N_R - 1$ (column three), given the degrees of polarization specified in the different rows.

<table>
<thead>
<tr>
<th>$(t_L, t_R)$</th>
<th>$N_L = N_R$</th>
<th>$N_L = 2 N_R - 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.45, 0.55)</td>
<td>31, 31</td>
<td>29, 15</td>
</tr>
<tr>
<td>(0.4, 0.6)</td>
<td>23, 23</td>
<td>25, 13</td>
</tr>
<tr>
<td>(0.3, 0.7)</td>
<td>13, 13</td>
<td>17, 9</td>
</tr>
<tr>
<td>(0.2, 0.8)</td>
<td>7, 7</td>
<td>9, 5</td>
</tr>
<tr>
<td>(0.1, 0.9)</td>
<td>3, 3</td>
<td>5, 3</td>
</tr>
</tbody>
</table>

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 $G_C$: 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 $G_C$ 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 $\Omega$ 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 $G$ 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 1 and proposition 2 apply to the closed membership game $G_C$ 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 $G_C$ 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 $G_C$ 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 7 states this result formally.

**Proposition 7** Let \(N = L \cup M \cup R\). 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 \(G_L\) 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 \(G\) 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 \(G\) 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 2 shows the necessary and sufficient condition for a voting bloc with a majority rule to be Individual-Exit stable. Propositions 5 and 7 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 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 \(N\) 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 \(r_N\) which may differ from simple majority, such that the (still exogenous) policy proposal passes if it gathers at least \(r_N N\) votes and a status quo stays in place otherwise.

The probability distribution over preference profiles is as before \(\Omega\), but types need not be independent. Rather, the only restriction that I impose for some results below is that \(\Omega\) 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 4** A voting bloc structure $(\pi, r)$ 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 $(\pi, r)$, 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 $(\pi, r)$.

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 8** Given a voting bloc structure $(\pi, r)$, 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 $(\pi, r)$ take place by $(\pi', r')$. Then $\sum_{i \in C'_{J+1}} u_i(\pi', r') \geq \sum_{i \in C'_{J+1}} u_i(\pi, r)$. 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 [11].

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 $r_N$ of the assembly.

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

**Proposition 9** Suppose $\Omega$ has full support. Let $(\pi, r)$ be any voting bloc structure. Let $(\pi', r')$ 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 $r_N = 1$, then $u_i(\pi', r') > u_i(\pi, r)$ for any $i$ in the new set of singletons $C'_0$ and $\sum_{i \in N} u_i(\pi', r') > \sum_{i \in N} u_i(\pi, r)$.

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 [11] 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 8 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 5 A voting bloc structure \((\pi, r)\) is Individual-Exit stable if 
\(u_i(\pi, r) \geq u_i(\pi', r)\) for any \(i \in N\) and any partition \(\pi' = \{C'_j\}_{j=0}^J\) such that:

(i) \(l \in C'_j \iff l \in C_j \) for all \(l \in N \setminus i\) 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 [14].

Definition 6 A voting bloc structure \((\pi, r)\) is Individually stable if 
\(u_i(\pi, r) \geq u_i(\pi_0, r)\) for any \(i \in N\) and any partition \(\pi_0 = \{C_0^j\}_{j=0}^J\) such that:

(i) \(l \in C_0^j \iff l \in C_j\) for all \(l \in N \setminus i\) and all \(j \in \{0, 1, ..., J\}\), and

(ii) for \(j \in \{1, ..., J\}\), if \(i \in C_0^j\) then \(u_l(\pi_0, r) \geq u_l(\pi, r)\) 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 7 A voting bloc structure \((\pi, r)\) is Nash stable if 
\(u_i(\pi, r) \geq u_i(\pi', r)\) for any \(i \in N\) and any partition \(\pi' = \{C'_j\}_{j=0}^J\) such that:

\(l \in C'_j \iff l \in C_j\) for all \(l \in N \setminus i\) 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 8 Let \((\pi^0, r^0)\) be the voting bloc structure in which all agents remain independent. A voting bloc structure \((\pi, r)\) is relevant if with positive probability the policy outcome under the structure \((\pi, r)\) differs from the outcome under \((\pi^0, r^0)\).

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 9** Let \((\pi, r)\) be a voting bloc structure with \(J\) blocs \(j = \{0, 1, \ldots, \hat{j}, \ldots, J\}\) such that \(r_0 = 1\). Let \((\pi, r')\) be such that \(r'_j = 1\) and \(r'_j = r_j\) for all \(j \in \{0, \ldots, J\}\)\(\setminus\hat{j}\). The voting bloc \(V_j = \{C_j, r_j\}\) is relevant in the structure \((\pi, r)\) if with positive probability the outcome under the structure \((\pi, r)\) differs from the outcome under \((\pi, r')\).

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

**Proposition 10** Suppose \(\Omega\) has full support and \(r_N \leq \frac{N-1}{N}\). 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 \(N_C \geq r_N N + 1\) and \(r_C N_C < r_N N\) 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 11** Suppose \(\Omega\) has full support and \(r_N \leq \frac{N-1}{N}\). Then there exists a relevant Individually stable voting bloc structure. If \(r_N \in \left(\frac{N+1}{2N}, \frac{N-1}{N}\right]\), then a voting bloc structure with a single bloc \((C, r_C)\) such that \(C = N\) and \(r_C < r_N\) 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 10, 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 12** Suppose \(\Omega\) has full support and \(r_N \in \left(\frac{N+1}{2N}, \frac{N-1}{N}\right]\). 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. \(C = N\), \(r_C < r_N\) is relevant and Nash stable.

If the voting rule of the assembly is simple majority, then proposition 11 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 voting bloc structure in which at least three agents form a voting bloc, and the following one describes characteristics of relevant Nash stable structures, provided they exist.
Proposition 13  A Nash stable voting bloc structure with at least one bloc of size at least three exists.

The grand coalition $C = \mathcal{N}$ with $r_C \in [r_N, 1]$ is irrelevant, but Nash stable. In a bloc $(\mathcal{N}, r_C)$ with $r_C \geq r_N$, 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 13 relates closely to corollary 12: 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, relevant Nash stable voting blocs have to be of size smaller than minimal winning.

Proposition 14  Suppose $\Omega$ has full support and $r_N = \frac{N+1}{2N}$. Then in any relevant Nash stable voting bloc structure $(\pi, r)$, $N_C < \frac{N+1}{2}$ for any voting bloc $\{C, r_C\}$ with a simple majority internal voting rule, and if there exist at least one singleton in $(\pi, r)$, then $N_C < \frac{N+1}{2}$ for any relevant voting bloc $\{C, r_C\}$.

Proposition 14 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. 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. If no single bloc is large enough to act as a dictator, each voting bloc affects the outcome with positive probability, and agents have incentives to coordinate in voting blocs.

The exact configuration into voting blocs in a stable voting bloc structure in the assembly depends on the size of the assembly and the specific probability distribution over preference profiles of the members of the assembly.

5 Conclusion and Extensions

Members of a democratic assembly -legislature, council, committee- 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.

The strategic formation of endogenous voting blocs has relevant implications for mechanism design and the selection of aggregation rules. From the perspective of a given coalition, such as a party or a confederation of nations trying to maximize the aggregate utility of its members subject to the constraint that no member wants to leave, I have shown that the optimal rule among the class of majority rules is the lowest supermajority such that every member of the coalition has an incentive to remain in the bloc. From the perspective of a social planner intending to maximize the utilitarian aggregate social welfare, the optimal rule for the assembly in the absence of voting blocs is a simple majority rule, but if agents coalesce strategically, the social planner must assess the merits of each rule based on the endogenous voting blocs that are expected to form with a given rule and simple majority may no longer maximize social welfare.

The theory in this paper has multiple natural extensions: Studying the robustness of the existence results to coalitional deviations using concepts such as Coalition-Proofness (see Bernheim, Peleg and Whinston [5]) or the Equilibrium Binding Agreements by Ray and Vohra [26]; 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.

6 Appendix

6.1 Proof of Lemma 1

To prove this lemma I first prove three intermediate steps. First I show that if a coalition $C$ leans right and the type of one of the members of $C$ shifts to the right (becomes higher), then the resulting coalition also leans right. Second I show that given a coalition $C$ that leads right, if the left-most member of $C$ (the member with the lowest type) leaves the coalition, then the resulting coalition also leans right. Third, I show that if $M$ is symmetric and $L$ leans left and forms a voting bloc, then the distribution of the number of votes cast by $L \cup M$ 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. Readers who find these three claims obviously true may skip down to the proof of Lemma 1.
Proof. Consider $C \setminus l$ and let $g_{C \setminus l}$ be the distribution function of $\sum_{i \in C \setminus l} s_i$. Its probability mass function $g_{C \setminus l}$ is determined by the aggregation of independent Bernoulli experiments, hence it is unimodal, as shown by Darroch [12]. Let $y$ denote the mode. Since preferences are independent, for any number $x$ and any agent $l$,

$$g^C(x) = t_l g_{C \setminus l}(x - 1) + (1 - t_l) g_{C \setminus l}(x)$$

Hence

$$g^C(x) - g^C(x) = (t_l - t_l)[g_{C \setminus l}(x - 1) - g_{C \setminus l}(x)]$$

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

For $k$ such that $N_C + k > y^{C'} \geq y$, it follows that

$$g^{C'}(N_C - k) \leq g^C(N_C - k) \leq g^C(N_C + k) \leq g^{C'}(N_C + k),$$

where the first and third inequalities hold by the sign of expression (1) and the second inequality by assumption.

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

The proof is similar for $C$ of odd size, with $g^{C'}(N_C - k) \leq g^{C'}(N_C + 1 + k)$ for any non-negative integer $k$.

Claim 16 Let $C \subseteq N$ be such that $N_C$ is even and let $t_i \leq t_i$ for any $i \in C$. Suppose that $g^C(N_C - 1 - k) \leq g^C(N_C + k)$ for all positive $k$. Then $g_{C \setminus l}(N_C - 1 - k) \leq g_{C \setminus l}(N_C + k)$ 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 $C' = \{C \cup l\} \setminus l$ with $t_l = 1/2$. By Claim 15, for any non-negative integer $k$, coalition $C'$ satisfies

$$g^{C'}(N_C - k) \leq g^{C'}(N_C + k).$$

The rest of the proof proceeds by induction. First, for $k = N_C - 1$, I prove that $g_{C'\setminus l}(0) \leq g_{C'\setminus l}(N_C - 1)$. 

26
To inequality (2), \( g^C(0) \leq g^C(N_C) \). It follows \((1 - t')g_{l,t'}(0) \leq t'g_{l,t'}(N_C - 1) \). Since \( t' = 1/2 \), then \( g_{l,t'}(0) \leq g_{l,t'}(N_C - 1) \).

Suppose that \( g_{l,t'}(N_C - 2 - k) \leq g_{l,t'}(N_C + k) \) holds for \( k = k' - 1 \). Then:

\[
t_{l,t'}g_{l,t'}(N_C - 2 - k') + (1 - t')g_{l,t'}(N_C - 2 - k') \leq t_{l,t'}g_{l,t'}(N_C + k - 1) + (1 - t')g_{l,t'}(N_C + k')
\]

Expression (3) implies (4) because \( t_{l,t'} = (1 - t_{l,t'}) = 1/2 \). Expression (4) implies (5) because \( g_{l,t'}(N_C - 2 - k') \leq g_{l,t'}(N_C + k') \) by assumption. Expression (6) is merely a reformulation of (5).

The induction argument is then complete. \( \blacksquare \)

**Claim 17** Let \( N = L \cup M \cup R \). Suppose \( g^M(N_M - 1 - k) = g^M(N_M + 1 + k) \), \( g^L(N_L - 1 - k) \leq g^L(N_L + 1 + k) \) 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.** Let’s 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

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

I first want to show that

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

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

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

\[
27
\]
for any $k$, rewrite inequality 7 as:

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

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

Regrouping terms:

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

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

$$P\left[\sum_{i \in L \cup M} v_i = \frac{N_L + N_M}{2} - k|L \text{ inactive}\right] \geq P\left[\sum_{i \in L \cup M} v_i = \frac{N_L + N_M}{2} + k|L \text{ inactive}\right]. \quad (8)$$

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

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

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

$$+ \sum_{h=0}^{r_L N_L - \frac{N_L + N_M}{2}} \left\{ P\left[\sum_{i \in L} s_i = \frac{N_L - 1}{2} - h|L \text{ inactive}\right]\left(P\left[\sum_{i \in M} s_i = \frac{N_M + 1}{2} + k + h\right] - P\left[\sum_{i \in M} s_i = \frac{N_M + 1}{2} - k - h\right]\right) \right\}.$$
**Proof.** For any \( h \in R \), let \( A_h \)

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

and similarly let \( B_h \)

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

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

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

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

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

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

\[
t_h \left[ \left( t_l P[ \sum_{i \in R_{-l}} s_i = r_R N_R - 2 ] + (1 - t_l) P[ \sum_{i \in R_{-l}} s_i = r_R N_R - 1 ] \right) P_1 \right] \]

\[
- \left( t_l P[ \sum_{i \in R_{-l}} s_i \leq (1 - r_R) N_R - 2 ] + (1 - t_l) P[ \sum_{i \in R_{-l}} s_i \leq (1 - r_R) N_R - 1 ] \right) P_3 \right]
\]

\[
+ (1 - t_h) \left[ \left( t_l P[ \sum_{i \in R_{-l}} s_i = (1 - r_R) N_R - 1 ] + (1 - t_l) P[ \sum_{i \in R_{-l}} s_i = (1 - r_R) N_R ] \right) P_2 \right] \]

\[
- \left( t_l P[ \sum_{i \in R_{-l}} s_i \geq r_R N_R - 1 ] + (1 - t_l) P[ \sum_{i \in R_{-l}} s_i \geq r_R N_R ] \right) P_4 \right],
\]

29
and \(t_l A_l + (1-t_l)B_l\) is equal to

\[
\begin{eqnarray*}
& & t_l \left[ 
\left( t_h P \left[ \sum_{i \in R_{-h}} s_i = r_R N_R - 2 \right] + (1-t_h) P \left[ \sum_{i \in R_{-h}} s_i = r_R N_R - 1 \right] \right) P_1 
\right. \\
& & - \left( t_h P \left[ \sum_{i \in R_{-h}} s_i \leq (1-r_R) N_R - 2 \right] + (1-t_h) P \left[ \sum_{i \in R_{-h}} s_i \leq (1-r_R) N_R - 1 \right] \right) P_3 \\
& & + \left( t_h P \left[ \sum_{i \in R_{-h}} s_i = (1-r_R) N_R - 1 \right] + (1-t_h) P \left[ \sum_{i \in R_{-h}} s_i \geq r_R N_R - 1 \right] \right) P_2 \\
& & \left. - \left( t_h P \left[ \sum_{i \in R_{-h}} s_i \geq r_R N_R - 1 \right] + (1-t_h) P \left[ \sum_{i \in R_{-h}} s_i \geq r_R N_R \right] \right) P_4 \right].
\end{eqnarray*}
\]

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

\[
(t_h - t_l) \left( 
\left( P \left[ \sum_{i \in R_{-h}} s_i = r_R N_R \right] + (1-r_R) N_R - 1 \right] P_1 - P \left[ \sum_{i \in R_{-h}} s_i = (1-r_R) N_R - 1 \right] P_2 \\
+ P \left[ \sum_{i \in R_{-h}} s_i \geq r_R N_R - 2 \right] - P \left[ \sum_{i \in R_{-h}} s_i \geq r_R N_R - 1 \right] \\
- \left( t_h P \left[ \sum_{i \in R_{-h}} s_i = (1-r_R) N_R - 1 \right] + (1-t_h) P \left[ \sum_{i \in R_{-h}} s_i \geq r_R N_R - 1 \right] \right) P_3 \\
\right).}
\]

Since \(M\) is symmetric and \(L\) leans left, it follows by Claim 17 that \(P_1 \geq P_2\) and \(P_3 \leq P_4\), and since \(R_{-h}\) leans right, by Claim 16 \(R_{-h}\) leans right as well. Then, the expression (11) above is weakly positive. ■

### 6.2 Proof of Proposition 2

**Proof.** By lemma 1, 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 1, \(l\) 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 \(t^{\text{Inc}}(r_R, r_L, t_{-l}) = -\frac{B_l}{A_l - B_l}\). Hence, it suffices to show that \(A_l \geq B_l\).

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

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

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

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

\[
30
\]
and
\[ P[\sum_{i \in R_{-i}} s_i \geq r_R N_R] \geq P[\sum_{i \in R_{-i}} s_i \leq (1 - r_R) N_R - 1]. \]

Since \( M \) is symmetric and \( L \) leans left,
\[ P[\sum_{m \in M \cup L} v_m \in \left[\frac{N_L + N_M}{2} - \frac{N_R - 1}{2}, \frac{N_L + N_M}{2} + \frac{N_R - 1}{2} - r_R N_R\right]] \geq P[\sum_{m \in M \cup L} v_m \in \left[\frac{N_L + N_M}{2} - \frac{N_R - 1}{2} + r_R N_R, \frac{N_L + N_M}{2} + \frac{N_R - 1}{2}\right]] \]
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, \( A_l \geq B_l \). ■

6.3 Proof of Proposition 4

This proof and the proof of proposition 7 use the result in proposition 8 below. While it is in principle inadvisable to use latter results in proofs that appear earlier in the text, proposition 8 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 4. In short, a voting bloc structure \((\pi, r)\) is a pair composed of a partition of the assembly \(\pi\), and a vector \(r\) that contains one rule for each voting bloc resulting from the partition \(\pi\).

**Proof.** Let \((\pi, r)\) be a voting bloc structure in which \(R\) does not form a voting bloc. Let \((\pi', r')\) 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 8, \(\sum_{i \in R} u_i(\pi', r') \geq \sum_{i \in R} u_i(\pi, r)\). Let \((\pi'', r'')\) 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 \((R \setminus i, r_R)\). 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 \(R \setminus i\) 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. ■

31
6.4 Proof of Proposition 5

Proof. From proposition 4, 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 \( P[\sum_{i \in M \cup L} v_i = \frac{N_R + 1}{2} - N_R] = \lambda \). 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 \( \sum_{i \in M \cup L} v_i = \frac{N_R}{2} - N_R \). 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 \( \varepsilon \) 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. \( \blacksquare \)

6.5 Proof of Proposition 6

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 \( \varepsilon A_h + (1 - \varepsilon)B_h > 0 \).

\[
A_h = P \left[ \sum_{i \in L_{-h}} s_i = r_L N_L - 1 \right] P \left[ \sum_{m \in M \cup L} v_m \in \left[ \frac{M + N_R}{2} - \frac{N_L - 1}{2}, \frac{M + N_R}{2} + \frac{N_L - 1}{2} - r_L N_L \right] \right]
- P \left[ \sum_{i \in L_{-h}} s_i \leq (1 - r_L) N_L - 1 \right] P \left[ \sum_{m \in M \cup R} v_m = \frac{M + N_R}{2} + \frac{N_L - 1}{2} \right].
\]

\[
B_h = P \left[ \sum_{i \in L_{-h}} s_i = (1 - r_L) N_L \right] P \left[ \sum_{m \in M \cup R} v_m \in \left[ \frac{M + N_R}{2} - \frac{N_L - 1}{2} + r_L N_L, \frac{M + N_R}{2} + \frac{N_L - 1}{2} \right] \right]
- P \left[ \sum_{i \in L_{-h}} s_i \geq r_L N_L \right] P \left[ \sum_{m \in M \cup R} v_m = \frac{M + N_R}{2} - \frac{N_L - 1}{2} \right].
\]

Let

\[
P_\alpha = P \left[ \sum_{m \in M \cup R} v_m \in \left[ \frac{M + N_R}{2} - \frac{N_L - 1}{2}, \frac{M + N_R}{2} + \frac{N_L - 1}{2} - r_L N_L \right] \right],
\]

32
\[ P_7 = P \left[ \sum_{m \in M \cup R} v_m \in \left[ \frac{M + N_R}{2} - \frac{N_L - 1}{2} + r_L N_L, \frac{M + N_R}{2} + \frac{N_L - 1}{2} \right] \right], \]

\[ P_6 = P \left[ \sum_{m \in M \cup R} v_m = \frac{M + N_R}{2} + \frac{N_L - 1}{2} \right] \]

Then,

\[ \varepsilon A_h + (1 - \varepsilon)B_h < \varepsilon \gamma \varepsilon^r L N_L^{-1} P_5 - \varepsilon (1 - \varepsilon)^{N_L - 1} P_6 + (1 - \varepsilon) \gamma \varepsilon^{(1-r)L} N_L^{-1} P_7. \]

Divide the right-hand side by \( \varepsilon \) and take the limit as \( \varepsilon \) 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 \( \varepsilon \) is low enough, \( \varepsilon A_h + (1 - \varepsilon)B_h < 0 \) and the voting bloc is not Individual-Exit stable.

\[ \square \]

6.6 Proof of Proposition 7

**Proof.** Let the vector of types be such that \( t_i = t_j \) for all \( i, j \in L \). Then, by proposition 8, 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 6, if the common type of \( L \) is low enough, \( (L, r_L) \) is not Individual-Exit stable.

\[ \square \]

6.7 Proof of Proposition 8

Proposition 8 and generalize proposition 1 in Eguia [15], from one to several voting blocs.

**Proof.** Let \( (\pi, r) \) be the initial voting bloc structure and \( (\pi', r') \) 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 \( \pi \) such that \( C'_{J+1} \cup C_0 = C_0 \) and \( C'_J = C_J \) for all \( j \in \{1, ..., J\} \). For notational simplicity, let \( C'_{J+1} \) be just \( C' \) and \( r'_{J+1} \) simply \( r'_C \).

Given \( s \), suppose \( \sum_{i \in C'} s_i = (1 - r'_C) N_{C'} \). Then \( \sum_{i \in C'} v_i(\pi', r', s) \leq \sum_{i \in C} v_i(\pi, r, s) \) since the votes of other agents are unaffected by the formation or not of a bloc \( (C', r'_C) \), it follows that \( \sum_{i \in N} v_i(\pi', r', s) \leq \sum_{i \in N} v_i(\pi, r, s) \). Hence, either the outcome is the same under \( (\pi, r) \) and \( (\pi', r') \), or if the outcome changes, it must be that the policy proposal passes under \( (\pi, r) \) but fails under \( (\pi', r') \) 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 \( N_{C'} - 2 \sum_{i \in C'} s_i \geq (2r'_C - 1) N_{C'} \geq 1. \)

Suppose instead that \( (1 - r'_C) N_{C'} < \sum_{i \in C'} s_i < r'_C N_{C'} \). 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 \( \sum_{i \in C'} s_i \geq r'_C N_{C'} \). 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. ■

6.8 Proof of Proposition 9

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 $(\pi', r')$, or it does not as in $(\pi, r)$. 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 $\Omega$ 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. ■

6.9 Proof of Proposition 10

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 $\sum_{i \in C} s_i = \sum_{i \in \mathcal{N}} s_i = r_CN_C < r_N\mathcal{N}$ 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, $\sum_{i \in C} v_i = \sum_{i \in \mathcal{N}} v_i = N_C \geq r_N\mathcal{N}$ and the proposal passes in the division of the assembly. Since $\Omega(s)$ has full support, $\sum_{i \in C} s_i = \sum_{i \in \mathcal{N}} s_i = r_CN_C$ occurs with positive probability and the bloc is relevant.

Since by assumption $N_C - 1 \geq r_N\mathcal{N}$, 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 $j \in C \setminus i$. 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 $r_N \leq \frac{N-1}{N}$ 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, $N_C \geq r_N N + 1$ and $r_C N_C < r_N N$ so that the second statement in the proposition applies. This is straightforward: If $r_N = \frac{N+1}{2N}$, let $N_C = N - 2$ and $r_C = \frac{N-1}{2(N-2)}$, and if $r_N > \frac{N+1}{2N}$, let $r_C = \frac{N+1}{2N}$ and $N_C = N$. ■

### 6.10 Proof of Proposition 11

**Proof.** First consider the with $r_N \in (\frac{N+1}{2N}, \frac{N-1}{N})$. Then, by proposition 10, any voting bloc structure with a unique voting bloc $(C, r_C)$ such that $C = N$ and $r_C < r_N$ 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 $r_N = \frac{N+1}{2N}$. Let $(\pi, r)$ 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 10, 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. ■

### 6.11 Proof of Proposition 13

**Proof.** The grand coalition $C = N$ with $r_C \in [r_N, 1]$ is irrelevant, but Nash stable. For any $s$ such that agent $v_i(N, r_C, s) = s_i$, agent $i$ is at least equally well off remaining in the bloc. For any $s$ such that $s_i = 0$ but $v_i(N, r_C, s) = 1$ it must be that $\sum_{j \in N \setminus i} s_j \geq r_C N \geq r_C(N - 1)$ 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 $v_i(N, r_C, s) = 0$ it must be that $\sum_{j \in N \setminus i} s_j \leq (1 - r_C)N - 1 \leq (1 - r_C)(N - 1)$ 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 $(N, r_C)$ with $r_C \in [r_N, 1]$ is

35
6.12 Proof of Proposition 14

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 $\sum_{h \in C} s_h \neq \frac{N_C}{2}$, it follows that $\sum_{h \in C} v_h \in \{0, N_C\}$ 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 $\sum_{h \in C} s_h = \frac{N_C}{2}$, any $i \notin C$ who joins the bloc causes $\sum_{h \in C} v_h = s_i N_C$ 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 $\sum_{h \in C} s_h = \sum_{h \in N} s_h = \frac{N_C + 1}{2}$. Then $\sum_{h \in C} v_h = \sum_{h \in N} v_h = N_C \geq \frac{N + 1}{2}$ and the proposal passes. If one of the non-members - who oppose the proposal - joins the bloc, then the expanded bloc is tied, $\sum_{i \in N} v_i = \sum_{h \in C \cup \{i\}} v_h = \sum_{h \in C} s_h < \frac{N + 1}{2}$ 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 $\Omega$ has full support, every preference profile occurs with positive probability and every non-member strictly prefers to join the bloc. Then, if $C \neq N$, the voting bloc structure is not Nash stable - a contradiction.

Suppose the voting bloc structure $(\pi, r)$ is such that $C_0 \neq \emptyset$, and $\exists (C, r_C)$ relevant such that $N_C \geq \frac{N + 1}{2}$. Let $(\pi', r)$ be a new voting bloc structure such that $C' = C \cup i$ and $C'_0 = C_0 \setminus i$ 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. ■
References


