Who Represents Whom: Strategic Voting and Conservatism in Legislative Elections*

Tilman Klumpp†
Emory University

September 2006

Abstract

This paper examines voting equilibria in a citizen-candidate model with multiple constituencies. Voters in the constituencies elect representatives into a legislative assembly to make a policy choice. Representatives both make policy proposals and vote on each others' proposals. The model thus formalizes the distinction between advocated policy and enacted policy in representation problems. Under the advocated policy aspect, constituents prefer representatives whose preferences are close to their own. Under the enacted policy aspect, constituents want representatives who prefer less change of the status quo than they do, as this helps to insure against extreme policy outcomes. If this second motive is strong enough, citizens elect conservative legislators who are relatively reluctant to change the status quo. We show that this happens when constituencies are sufficiently heterogeneous with respect to their policy preferences. Our results shed light on a number of political phenomena. Particular attention is devoted to the issue of reform obstacles. We show show that legislative resistance to reform can arise in equilibrium even for projects that enjoy broad popular support in the electorate. Unlike in existing models, however, the reform deadlock is incomplete, and some reform will be undertaken in equilibrium.

Keywords: Political Representation, Legislative Elections, Citizen-Candidate Model, Strategic Delegation, Conservatism.

JEL code: D72, D78.

---

*I thank Marco Battaglini, Massimo Morelli, Mattias Polborn, Eric Rasmusen, and audiences at Indiana University, the University of Alabama, Emory University, the 2003 Canadian Economic Theory Conference, and PET 2004.

†Department of Economics, Emory University, Rich Building 316, 1602 Fishburne Dr., Atlanta, GA 30322. E-mail: `tklumpp@emory.edu`.

1 Introduction

Policy making in modern democracies almost always involves large legislative bodies whose members are elected representatives of their constituencies. Political representation of citizens can ensure that local interests of constituents are taken into account when formulating policy. In addition, there are efficacy reasons for representative democracy: Only a few citizens, the legislators, are fully involved in the complicated and lengthy procedure of policy making. Since legislators are elected representatives of their constituents, however, citizens "have a voice" throughout the process. The legislative process in representative democracies is therefore hierarchical: First citizens choose their representatives through legislative elections. Second, the elected representatives choose policy by means of bargaining, voting, or some other form of compromise. In this process, representatives assume two roles. One aspect is that representatives can *advocate* policies for their constituents. In congressional election campaigns, for example, candidates often promise that "If elected, I will fight for legislation X ... ." This typically involves sponsorship of a particular bill, or serving on congressional subcommittees to draft legislation. A second aspect of representation is that representatives eventually have to *enact* some policy through their votes in the legislative assembly.[1] The incentives a representative faces across these two roles may not always coincide. For example, a representative may vote in favor of some proposed legislation, even though it does not reflect his particular views very well, as long as he prefers it to the status quo. In fact, another statement that is often made in congressional campaigns, typically by new entrants against incumbent representatives, is that "He promised to fight for X, but then voted for Y ... ." Such statements reflect not only a distinction between the policies a representative advocates and those he votes for, but also the inability of candidates for legislative office to commit to their votes once elected.

This paper examines legislative policy making in a model of hierarchical political representation. Citizens are grouped into constituencies and elect representatives into a legislative assembly. The policy space is one-dimensional and preferences are single-peaked. Once a legislature is determined, some representative is appointed agenda setter at random and given the right to make a policy proposal. All representatives then vote for or against the proposal, and depending on the outcome of this vote the proposal is either implemented or the status quo policy is maintained. Representatives are regular citizens and cannot make binding promises of how to act as legislators. A legislature is a *representation equilibrium* if no representative can be defeated in a pairwise election against another candidate in his constituency. The identity of the winner in one district will depend on the representatives elected in the other districts. Thus, when electing their representatives, voters in one constituency will take into account their expectations of who will be elected in the other constituencies, and these expectations will be correct in equilibrium.

We show that if constituencies are sufficiently similar to each other in terms of the distribution of preferences of their members, each constituency elects its median voter to the legislature. It is also possible, however, for citizens to elect delegates whose political stance is different from the median policy preference among the representative's constituents. This will happen if constituencies are sufficiently dissimilar, and in this case

[1] The terminology of "advocating" vs. "enacting" policy is borrowed from the political science literature; see Cox (1997).

delegates will have ideal points that are closer to the status quo than those of a majority of their constituents. We call legislatures and policy outcomes that are distorted in such a way *conservative*.[2] The reason why citizens may want to elect such conservative representatives is an insurance motive that is directly linked to the inability of representatives to commit to certain actions. Note that a representative will vote for those policy proposals he prefers to the status quo, and the closer his ideal point is to the status quo, the smaller is the set of proposals he approves. Consider now a citizen whose ideal policy outcome entails a moderate change the status quo. A conservative representative, i.e. one whose ideal point lies between the status quo and the citizen's ideal point, can thus credibly promise not to vote for legislation that overshoots the citizen's policy goal. If this representative is pivotal in the legislature, his conservative preferences insure the moderate citizen against extreme policy outcomes. A representative who shares the citizen's preferences, on the other hand, would vote for larger set of proposals and would thus not be able to provide the same degree of insurance. Of course, there is a cost in electing a conservative delegate, namely if this delegate is the agenda setter. In this case, the resulting legislation is also unsatisfactory from the citizen's perspective, only that now it falls short of the citizen's ideal outcome. Whether or not the citizen votes for such an unrepresentative representative therefore depends on whether the benefit of insuring against extreme policies outweighs this cost or not. We show that the former is the case whenever the degree of preference dispersion *across* constituencies is sufficiently high. Intuitively, if the dispersion of political views across constituencies becomes larger, the distribution of proposed policies exhibits greater variation. Once this variation becomes too great, moderate citizens would like to prevent policies which are too extreme from being implemented, and they can achieve this goal by electing conservative delegates. Thus, in our paper conservatism is the consequence of *strategic delegation* of political power from voters to representatives who do not share the the same preferences as their median constituents.[3]

We present sufficient conditions on the degree of heterogeneity for conservatism to arise in all or no equilibrium outcomes. For the special case of linear utility, we are able to provide exact bounds on the degrees of preference dispersion across districts beyond which conservative legislatures arise. Preference heterogeneity thus translates into political outcomes which are distorted relative to the myopic case in which constituents elect their median voters. These distortions can be strong enough to result in inefficient policies, meaning that a majority of citizens in every constituency can be made better off under a different legislature. This inefficiency is driven by an individually rational, but in the aggregate suboptimal, desire to insure against extreme policies. Furthermore, outcomes are not necessarily monotone in preferences. For instance, it is possible that if every citizen's ideal point shifts to the right, the policies enacted in equilibrium shift to the left.

[2] The term *conservative* is used here in the meaning of "being reluctant to change" and not in the sense of "holding rightist beliefs."

[3] It is important to point out that, while using the term "insurance" repeatedly, the paper's results do not depend on risk aversion (beyond the fact that single peaked utility functions are quasi-concave in the policy outcome). In particular, all of our results hold for the case of linear utility. The reason is that the cost of electing a conservative representative—a utility loss if that representative has agenda power—is often smaller than the benefit it brings—a gain if a more extreme representative has agenda power and is constrained. Conservatism therefore provides insurance at a "below-fair price," so even risk-neutral individuals may seek it.

This is the case if the preference shift is accompanied by a sufficiently large increase in preferences dispersion across constituencies.

These results shed light on several political phenomena which we discuss in some detail in Section 6 of the paper. First, our model can help to explain the apparent unrepresentativeness of U.S. Congress that has been documented in a number of studies. Second, the conservatism effect we find creates substantial inertia in the political process, making changes of the status quo more difficult (but not impossible). Thus our model delivers insights to the issue of reform resistance; in particular, it provides a novel theory of incomplete reform deadlocks and insufficient reform. Third, the paper provides a rational theory of so-called "backlash effects" observed by sociologists: As the representation of racial and ethnic minorities in U.S. state legislatures increased, representatives for the majority group have become less likely to sponsor legislation that advances minority interests. Numerous different explanations exists for these observations. There are certain features of the political and economic environment, however, which can be common factors facilitating each of these distortions. In particular, we argue that heterogeneity of preferences across constituencies might be one such factor, and understanding the link between preference heterogeneity, representation, and conservatism is crucial for understanding these distortive effects in representative democracies.

As a description of real policy making processes, our model of indirect democracy is of course unrealistic. However, it captures some general aspects which play an important role in many instances of political representation. The first aspect is the inability of representatives to commit to a certain behavior in the legislature. Our model is hence one of *citizen candidates*, and the representatives' preferences determine their actions once elected. The second aspect is that constituents not only care about their own representatives' role as policy advocates, but also about their abilities to constrain other representatives. The agenda setter framework we adopt is a simple, but useful model, in this respect. The possibility of being recognized as proposer in the legislature corresponds to a representative's advocacy role, while the right to vote for or against a particular proposal corresponds to the enactment role. Since the agenda setter is randomly selected in the legislature, there will be uncertainty regarding the final policy outcome. This uncertainty assumption is crucial, but well motivated: What we intend to capture is the idea that citizens cannot perfectly anticipate the policies that will emerge from a possibly diverse legislature over the duration of their term. For instance, proposed bills in U.S. congress are typically reported out of some congressional subcommittee, and the selection of these committees is governed by clear congressional rules. Nevertheless, from the viewpoint of a citizen who is not familiar with the details of these rules, there may well be uncertainty as to which delegate will be selected to serve on a particular committee, and how much influence he or she will have in drafting the proposal.[4] These features are common to legislative processes in many countries, regardless of the fine details of the political system under consideration, and our model, while stylized, reflects these general characteristic very well.

[4] An alternative justification for our modeling assumptions is that in a typical legislative term, many bills regarding different issues are introduced by different legislators. Suppose that a legislator's preferences on one issue are similar to that on other issue (see Poole and Rosenthal, 1991, for empirical evidence in support of such an assumption.) If each legislator makes proposals on the same number of issues, the multi-issue scenario is a replication of the simple model described above.

The remainder of the paper is organized as follows. In Section 3 we introduce the formal framework and define the concept of representation equilibrium. In Section 4 we present our main characterization and existence results. Section 5 links conservatism to inter-district heterogeneity. In Section 6, we use our model and its results to interpret various political phenomena. In Section 7 we examine coalition-proof equilibria. Section 8 concludes. All but very short proofs are in the Appendix.

2 Related Theoretical Literature

A relatively early literature on legislative elections utilizes the Downsian framework. Austen-Smith (1984) formulates a model where the final policy is set by party that wins control over the legislature; the party's policy in turn depends on its members' individual policy platforms on which they run in their constituencies. Like in our paper, this necessitates that citizens look past their own districts' candidates, and do not necessarily vote for the candidate whose individual stance is closest to their own. Austen-Smith (1984) shows that in such a framework, the median voter result is generally preserved for party policies, but not for the policies chosen by the individual candidates.[5]

Citizen-candidate models, as an alternative to the Downsian assumption, were introduced by Osborne and Slivinski (1996) and Besley and Coate (1997) for the case of a single district. Modeling candidates as citizens circumvents potential commitment problems in an elegant and convincing way: Since candidates are citizens with preferences of their own, voters know that elected officials will act in a way that advances their own preferences, subject only to whatever post-election constraints politicians face. Hence electing a politician boils down to a delegation problem, where a citizen votes for the candidate who, once elected, is most likely to act in a way aligned with the citizen's own objectives.[6]

Two-district citizen-candidate models are examined in Besley and Coate (2003) and Redoano and Scharf (2004). Policy choice in these papers concerns the provision of local public goods in the presence of spillovers, and while strategic delegation effects can occur in both papers under certain assumptions, they are very different from the delegation effect in our paper. Besley and Coate (2003) show that when policy-making among the legislators takes the form of joint surplus maximization, the election of representatives who favor a higher level of public goods than a majority of their constituents is possible. This effect is weaker the stronger the spillovers across constituencies. The opposite occurs in the model of Redoano and Scharf (2004). When the legislators choose district-specific public good levels, free-riding on the other district's public policy provision leads to the election of representatives who favor a lower level of public goods than a majority of their constituents. This effect is stronger the stronger the spillovers. In our agenda-setter model, on the other hand, representatives not only propose policy, but also vote on policies proposed by others, and it is the tradeoff between these two roles that gives rise to strategic delegation.

An N-district citizen-candidate model is Chari, Jones, and Marimon (1997). Like ours, their model uses a single-offer bargaining game to determine committee decisions.

[5]See Austen-Smith (1986) for an extension of the model that does not presume parties from the outset.

[6]That this person need not have one's own preferences has been demonstrated in other contexts of delegation; see Persson and Tabellini (1994), Rogoff (1985), Alesina and Grilli (1992), and Cai (2000).

However, the main focus of Chari et al. (1997) is on explaining the phenomenon of split-ticket voting, i.e. voting for political candidates with differing views in congressional and presidential elections. Our model abstracts from the executive branch and describes policy choice in a simple legislature. Further, Chari et al. (1997) consider a distributional setting and hence multi-dimensional policies (the case of a "global policy variable," such as foreign policy, etc., is treated only peripherally there.) While there is a strategic delegation effect, it is again much different from ours. Equilibrium representatives in Chari et al. (1997) will be "big spender" types who prefer a higher level of public expenditures than their constituents. This is a consequence of the fact that in a model with targeted benefits, the agenda setter will secure a minimal winning coalition for his proposal, and representatives tolerant of government spending are more likely to be included in such coalitions. Overall spending may be inefficient as a result, and distorted away from the status quo of zero expenditures. The goal for districts in our paper is not to be included in a minimal winning coalition; in fact, it is possible that an equilibrium proposal receives the support of more legislators than needed for a majority. Instead, since some districts want to reduce the risk of extreme outcomes by electing conservative delegates, distortions are toward the status quo. Finally, by assuming identical preference distributions across constituencies, the issue of heterogeneity does not arise in their paper.[7]

Morelli (2004) studies a three-district model of legislative elections that lies between the Downsian and the citizen-candidate framework. Candidates are citizens with preferences of their own; however, they can offer other platforms provided they belong to a party that commits them to such a platform. The number of effective parties is endogenous in the model, since two smaller parties can merge before the election to form a larger and broader party. The main question the paper addresses concerns the number of effective parties in different electoral systems (i.e. the Duvergerian prediction), and is thus quite different from the issues we address in this paper. The main result is that the Duvergerian prediction can be reversed (i.e. there are fewer mergers and hence more parties under plurality rule than under proportional representation) if preferences across districts are dissimilar. This dependence of the number of parties on the distribution of preferences across districts bears a resemblance to the conservatism effect in our paper, and like our result it is related to the bargaining aspect of the model. However, bargaining in Morelli (2004) concerns pre-election compromise among party leaders, and not post-election compromise among legislators as in our paper, and is hence not comparable to our model.

Finally, with regard to legislative bargaining games, a growing literature examines collective choice in an exogenously given committee. These papers do not consider the formation of this committee, and mostly focus on settings where policy concerns the distribution of benefits to legislators and their constituencies. The seminal paper to contain a formal model of the legislative bargaining process is Baron and Ferejohn (1989), which contains both a simple two-round bargaining game over distributions, and a more complex repeated game.[8] A departure from the distributional setting is Banks and Duggan (2000),

[7]Other models which distinguish between the executive and legislative branch of government are found in Alesina and Rosenthal (1996), and Persson, Roland, and Tabellini (1998, 2000).

[8]Related models that also use an infinitely repeated process are Eraslan (2002), Eraslan and Merlo (2002), and Morelli (1999). Papers that study finitely repeated bargaining include Battaglini and Coate (2005), and Bernheim, Rangel, and Rayo (2004).

who examine infinitely repeated bargaining in a spatial model and prove a median voter result. Our model contains a much simpler approach to the bargaining process than these papers, and focuses instead on the composition of the legislative body which is determined endogenously in equilibrium.

3 The Model

3.1 Citizens, policies, and delegates

The policy space is the real line $\mathbb{R}$, and a typical policy is denoted by $x \in \mathbb{R}$. The default policy, or status quo, that is in place before the political process is started is x_0, which we normalize to $x_0 = 0$ without loss of generality. All citizens have single peaked preferences over policy outcomes: If a citizen has ideal policy point $\psi \in \mathbb{R}$ and policy $x \in \mathbb{R}$ is enacted, she obtains utility $u(d)$, where $d = |x - \psi|$ is the distance between the citizen's bliss point and the policy outcome. We assume that $u : [0, \infty) \rightarrow \mathbb{R}$ is continuous and twice differentiable, strictly decreasing on $\mathbb{R}_+$ ($u'(d) < 0$ for $d > 0$), and concave ($u''(d) \leq 0$). Since utility is defined as a function of the distance between policy outcomes and bliss points, it must necessarily be symmetric over policies, i.e. the policies $x = \psi - d$ and $x' = \psi + d$ give the same utility to a citizen with bliss point ψ. Special cases of such preferences include linear utility ($u(d) = -d$), and quadratic utility ($u(d) = -d^2$).

The set of citizens is partitioned into N electoral districts, or constituencies, indexed $i \in I = \{1, \ldots, N\}$. We assume that N is odd for analytical convenience.[9] We define $m \equiv \frac{N+1}{2}$, the simple majority of N. In each district resides a continuum of citizens whose bliss points are distributed on $\mathbb{R}$ with positive density everywhere. Denote by ψ_i the median bliss point in district i. Again without loss of generality, we assume that districts are ordered by their median bliss points, that is, $\psi_1 < \psi_2 < \ldots < \psi_N$, and that $\psi_m > 0$.

A legislator for district i is a citizen of i.[10] Let $\phi_i \in \mathbb{R}$ be the bliss point of the representative for district i, and call the vector $\phi = (\phi_1, \ldots, \phi_N) \in \mathbb{R}^N$ an *assembly*. We use notation $\phi^{(m)}$ to denote the median legislator in ϕ (i.e. the m-th highest entry in the vector ϕ).

Since enacted policy is simply a change of the status quo, the underlying preferences over policies can be regarded as preferences over policy reforms. This interpretation motivates a notion of conservatism that has the interpretation of being reluctant to change: Person A is conservative relative to person B if A prefers a smaller magnitude of change of the status quo than B does. One can also call such a reluctance to change *structural conservatism*, in order to distinguish it from other notions of conservatism such as social, religious, or fiscal conservatism. In our model, a representative with bliss point ϕ_i is conservative relative to a constituent with bliss point ψ if $|\phi_i| < |\psi|$. Note that a representative who is conservative relative to the median voter of his constituency is necessarily conservative relative to a majority of his constituents. Accordingly, we may define a conservative legislature as follows:

[9]Qualitatively our results would be unaffected if N was even; however allowing for general N would complicate the notation.

[10]We use the terms representative, legislator, and delegate interchangeably.

Definition 1. An assembly ϕ *exhibits conservatism* if $|\phi_i| \leq |\psi_i| \; \forall i \in I$, with at least one inequality strict.

We now turn to the political process: First, citizens elect legislators as their representatives by majority voting. Then the elected representatives produce a collective policy decision via a simple game. We start with the second phase.

3.2 The legislative committee decision

To model legislative decision making, we adopt the random proposer model of Romer and Rosenthal (1978). Given an assembly ϕ, legislation is initiated by an agenda setter who is selected randomly among the N legislators. The probability that any given legislator is the agenda setter is $\frac{1}{N}$, and we let $A \in I$ denote the agenda setter's identity. Legislator A makes a policy proposal x on which the assembly then votes. Letting $v_\phi(x)$ be the number of votes in favor of adoption of x, the proposal x is implemented if and only if $v_\phi(x) \geq m$; otherwise the status quo policy $x_0 = 0$ is maintained.

We assume that a legislator who is indifferent between the status quo and the new proposal votes for the latter. Faced with proposal x, representative i's vote v_i is given by

$$v_i(x) \;=\; \begin{cases} 1 & \text{if } u(|x-\phi_i|) \geq u(|0-\phi_i|), \\ 0 & \text{otherwise,} \end{cases} \;=\; \begin{cases} 1 & \text{if } |x-\phi_i| \leq |\phi_i|, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where $v_i(x) = 1$ (0) means that legislator i votes for (against) the proposal x. For a proposal to win, it needs the support by a majority of legislators:

$$v_\phi(x) \equiv \sum_{i \in I} v_i(x) \geq m, \tag{2}$$

which is the case if and only if the bliss points of at least m members of the legislature are at least as close to x as they are to zero. Define

$$Q(\phi) = \{x \in \mathbb{R} : v_\phi(x) \geq m\} \tag{3}$$

to be the set of proposals that win against the status quo. Observe that a representative with $\phi_i \leq 0$ votes for all proposals $x \in [2\phi_i, 0]$, and a representative with $\phi_i \geq 0$ votes for all proposals $x \in [0, 2\phi_i]$. Thus, the set $Q(\phi)$ is a closed interval, given by

$$Q(\phi) = \begin{cases} [0, 2\phi^{(m)}] & \text{if } \phi^{(m)} \geq 0, \\ [2\phi^{(m)}, 0] & \text{if } \phi^{(m)} < 0. \end{cases}$$

We let $\overline{Q}(\phi) \equiv \max Q(\phi)$ and $\underline{Q}(\phi) \equiv \min Q(\phi)$. Note that $0 \in Q(\phi)$ for all ϕ, so A can always guarantee herself utility $u(|\phi_A|)$. To maximize her utility, A makes the proposal that is closest to ϕ_A and still draws a majority of votes,

$$z_A(\phi) = \text{arg min}_{x \in Q(\phi)} |\phi_A - x|. \tag{4}$$

The proposal function z is well defined, since we assumed that a representative who is indifferent between the status quo and x votes for x; hence $Q(\phi)$ is closed. Note further that $z_A(\phi)$ weakly increases in ϕ. That is, given $\phi, \phi' \in \mathbb{R}^N$ with $\phi \geq \phi'$, we have $\underline{Q}(\phi) \geq \underline{Q}(\phi')$ as well as $\overline{Q}(\phi) \geq \overline{Q}(\phi')$, so that shifting ϕ weakly to the right shifts $Q(\phi)$ weakly to the right. Consequently, (4) implies that z_A is weakly increasing in ϕ.

The agenda setter's proposal is a function of ϕ because $Q(\phi)$ determines what policies can be implemented. This constraint varies with ϕ, so A may make different proposals to different assemblies. If $z_A(\phi) = \phi_A$, we say that A is unconstrained. If $z_A(\phi) \neq \phi_A$, the agenda setter is constrained in that she cannot implement her most preferred policy. In any case, A never makes a proposal if she anticipates rejection. Once ϕ and A are determined, $x^* = z_A(\phi)$ is the final policy outcome. Given ϕ, but before the agenda-setter is selected, the expected policy outcome is $E(x^*|\phi) = \frac{1}{N}\sum_{i \in I} z_i(\phi)$.

3.3 Voting in the constituencies

At the first stage of the democratic process, citizens in each electoral district i choose a delegate to represent them in the legislature. We assume that representatives who are selected are Condorcet winning candidates in each district.[11] To break ties, we assume that if a voter is indifferent between two candidates which will propose the same policy to the legislature, she votes for the one whose bliss point is closer to the voter's bliss point.[12] Formally, let $U_\psi(\phi)$ the expected utility a citizen with bliss point ψ obtains if ϕ is the elected assembly (we will examine this function in more detail in the next section). Then we assume the following:

Assumption 1. Consider a citizen in constituency i with bliss point ψ, a set of representative ϕ_{-i} for the other constituencies, and two candidates ϕ_i, ϕ_i' in a pairwise election in district i such that $z_i(\phi_{-i}, \phi_i) = z_i(\phi_{-i}, \phi_i')$, $U_\psi(\phi_{-i}, \phi_i) = U_\psi(\phi_{-i}, \phi_i')$, and $|\psi - \phi_i| < |\psi - \phi_i'|$. Then the citizen with bliss point ψ votes for candidate ϕ_i.

All district-level elections are held simultaneously. Formally, we define a political equilibrium as follows:

Definition 2. A *representation equilibrium* is an assembly $\phi^* = (\phi_1^*, \ldots, \phi_N^*)$ such that for every $i \in I$ the following holds: Given ϕ_{-i}^*, ϕ_i^* cannot be defeated through majority voting in district i by some $\phi_i \neq \phi_i^*$.

Our definition of equilibrium requires that among the members of i, there exists one with bliss point ϕ_i^*. Since we have assumed that districts are populated by continuum of voters with positive density everywhere, this requirement is met. The continuum assumption is of course unrealistic, but mild as long as the number of voters is large and voters are sufficiently diverse. Then, even if an "ideal representative" ϕ_i^* is not available in district i, there should be one who is very close to ϕ_i^*.

4 Representation Equilibrium

This section examines equilibrium voting behavior in the first stage of the two-stage game described above. First, we show that any assembly which satisfies Definition 2 of equi-

[11] We prove in Section 4.1 that such a delegate always exists.

[12] In terms of policy outcomes, this assumption does not change our results. It is only made to avoid multiple equilibria that are policy-equivalent.

librium is a pure strategy Nash equilibrium of a related game, played only between the median voters of the N constituencies. This related game is a *delegation game*, in which each player chooses a person to make decisions for him. Next, we look at the incentives a single player faces in this delegation game. In particular we demonstrate in a simple three-district example that it can be optimal to given decision making power to a person with preferences different from one's own. Last, we show how this incentive translates to equilibrium outcomes. We prove the existence of representation equilibria and characterize a certain class of them in terms of their degree of conservatism. We also present several further examples.

4.1 Legislative elections as a delegation game

Recall from the previous section our assumption that a citizen's utility is decreasing in the distance between enacted policy and his bliss point, which means that voters have single peaked preferences over policies. Hence, if policy was chosen in a direct democracy, the *overall* median voter's most preferred policy would be the Condorcet winning policy, as predicted by the median voter theorem. In our model of indirect democracy, however, citizens in each constituency must elect a representative, and this representative then bargains over policy with the representatives from other constituencies. Electing a legislator is therefore equivalent to choosing an agent to play the committee game for his constituency. While citizens possess underlying preferences over enacted policy, through the legislative process they have induced preferences over legislators. We now examine these induced preferences more closely.

Given some vector $\phi \in \mathbb{R}^N$ assume that $\phi^{(m)} \geq 0$, so $Q(\phi) = [0, 2\phi^{(m)}]$. The expected utility of assembly ϕ for a citizen with bliss point ψ can be written as follows:

$$\begin{aligned} U_\psi(\phi) &= \frac{1}{N} \sum_{i \in I} u(|z_i(\phi) - \psi|) \\ &= \frac{1}{N} \left[\sum_{i:\phi_i < 0} u(|\psi|) + \sum_{i:\phi_i \in Q(\phi)} u(|\phi_i - \psi|) + \sum_{i:\phi_i > 2\phi^{(m)}} u(|2\phi^{(m)} - \psi|) \right]. \end{aligned} \quad (5)$$

If $\phi^{(m)} < 0$, a similar expression can be obtained. In the second line of (5), the first term represents the utility that the citizen obtains if a representative to the left of the constraint set is agenda setter, the second term represents the utility if an unconstrained representative is agenda setter, and the third term the utility if a representative to the right of the constraint set is the agenda setter. Notice that the location of the bliss point of i's delegate, ϕ_i, has two effects on final policy outcomes. If i is agenda setter, she proposes the policy that is closest to ϕ_i among those policies that will draw a majority of votes. Therefore, when i is the agenda setter ($A = i$), ϕ_i enters the proposal function z_i in (4) in the $|\phi_A - x|$ term. However, if some other representative j is agenda setter ($A = j \neq i$), i's vote may be necessary for j's proposal to pass. In this case ϕ_i enters the proposal function as part of the constraint set $Q(\phi)$.

Holding ϕ_{-i} fixed, $U_\psi(\phi_{-i}, \phi_i)$ may peak at more than point $\phi_i \in \mathbb{R}$, so preferences over legislators have a different shape than underlying policy preferences. In particular, single-peakedness of policy preferences does not imply the existence of a Condorcet winner in each district. Nevertheless, U_ψ satisfies the following single-crossing property, introduced by Gans and Smart (1996):[13]

Lemma 1. *Given $\phi, \phi' \in \mathbb{R}^N$ with $\phi \geq \phi'$, there exists $\hat{\psi} \in \mathbb{R}$ such that $U_\psi(\phi) \leq U_\psi(\phi')$ for all $\psi \leq \hat{\psi}$, and $U_\psi(\phi) \geq U_\psi(\phi')$ for all $\psi \geq \hat{\psi}$.*

The proof of Lemma 1 is in the Appendix. The result guarantees the existence of a Condorcet winner in each legislative district, as the following Theorem states:

Theorem 1. *Given ϕ_{-i}, there exists ϕ_i^* that cannot be defeated through majority voting in district i by some $\phi_i \neq \phi_i^*$. Furthermore, let $\Phi_i^* = arg\ max_{\phi_i \in \mathbb{R}} U_{\psi_m}(\phi_{-i}, \phi_i)$; then the winning candidate in district i is $\phi_i^* \in arg\ min_{\phi_i \in \Phi_i^*} |\psi_i - \phi_i|$.*

Proof. In the absence of Assumption 1, Φ_i^* is the set of Condorcet-winning candidates in district i, given ϕ_{-i}. Given Lemma 1, $\Phi_i^* \neq \emptyset$ (for a proof of this result see Gans and Smart, 1996). Using Assumption 1, we have $\phi_i^* \in \arg\min_{\phi_i \in \Phi_i^*} |\psi_i - \phi_i|$ as stated. □

Theorem 1 alone does not imply that a political *equilibrium* exists (we will prove existence below). However, the result allows us to treat voting in the constituencies as a noncooperative game played among the median voters $\psi_1, \ldots, \psi_N$. In this game, player $i = 1, \ldots, N$ selects a strategy $\phi_i \in \mathbb{R}$ and obtains payoff

$$U_i(\phi) \equiv U_{\psi_i}(\phi_{-i}, \phi_i)$$

from the strategy profile $\phi = (\phi_1, \ldots, \phi_N)$. A pure strategy Nash equilibrium of this dual game, if it exists and satisfies the same tie-breaking assumption as the original voting game, will be an equilibrium legislature according to Definition 2. Hence all that matters for political outcomes is the location of a district's median voter, ψ_i. In the following, therefore, a district is completely described by the value ψ_i, and no further information about the distribution of preferences around the median is required.

4.2 Incentives in the delegation game: An example

The following example illustrates the insurance effect that can lead to conservatism in legislative elections. Consider $N = 3$, and suppose $\psi_1 < 0 < \psi_2 < \psi_3$. Suppose first that each district elects its median voter as its representative, so that $\phi_i = \psi_i$, $i = 1, 2, 3$. Figure 1 depicts this case. The three representatives' bliss points are located on the horizontal line, and the three boxes above are the proposals which each of the representatives will approve against the status quo policy, which is zero. In order for a proposal to be implemented it needs the votes of at least two representatives, and the shaded region indicates for which proposals this will be the case; i.e. the shaded region represents the set $Q(\phi)$. The only representative who is constrained is ϕ_1; if he has agenda power he does not propose his most preferred point, but instead proposes $z_1(\phi) = 0$. Both ϕ_2 and ϕ_3, however, are within the set $Q(\phi)$, so $z_i(\phi) = \phi_i = \psi_i$, $i = 2, 3$. Since each representative has a one-third chance of being recognized, the distribution over possible policy outcomes is uniform over the set $\{z_i(\phi)\}_{i=1,2,3}$; these points are indicated at the bottom of the graph.

Now consider a scenario where the median district elects a conservative representative, $|\phi_2| < |\psi_2|$, as depicted in Figure 2.

[13]This property is also known as *order-restricted preferences*.

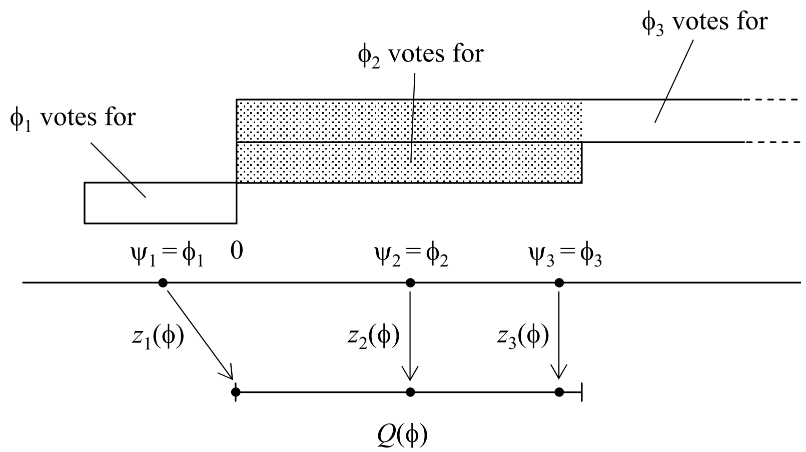

Figure 1: Policy outcomes in a “truly representative” legislature

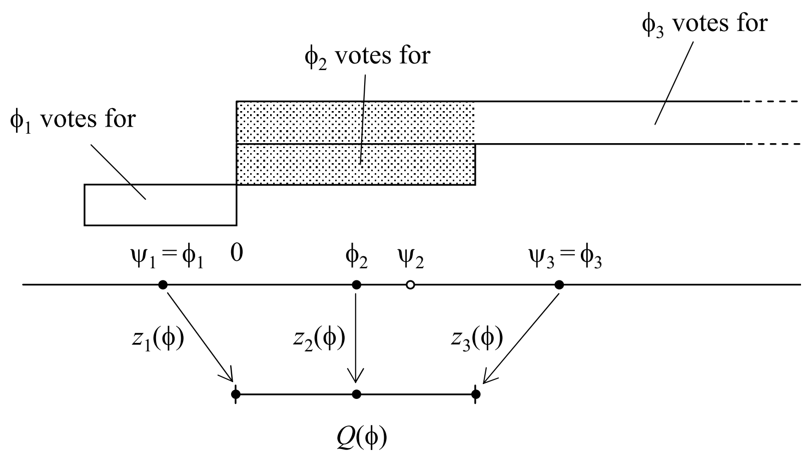

Figure 2: Policy outcomes in a conservative legislature

Note that in both Figure 1 and 2, the representative from district 2 effectively controls the set $Q(\phi)$ of implementable policies. With a conservative representative, $Q(\phi)$ shrinks. In particular, in Figure 2 the set $Q(\phi)$ is small enough to constrain not only the representative from district 1, but also the representative from district 3 who now makes the proposal $z_3(\phi) < \psi_3$. This new proposal is closer to ψ_2 and is thus preferred by the median voter in district 2. However, shrinking the set $Q(\phi)$ in this way has a cost, namely that the policy proposal made by district 2's representative is farther away from ψ_m than before. Overall, it is possible that the benefit of constraining ϕ_3 outweighs the cost of having a conservative ϕ_2. In Figure 2, the movement of $z_2(\phi)$ away from ψ_2 is smaller than the movement of $z_3(\phi)$ toward ψ_2. So with linear utility, for example, the expected utility under the conservative assembly is higher for ψ_2 than under the “truly representative” assembly.

4.3 Existence and characterization of equilibrium

While the previous example illustrates the reason why it might be desirable to be represented by a person with different political views, it does not tell us what we can expect as equilibrium outcomes of the delegation game. We now turn to the formation of equilibrium legislatures. In order to characterize representation equilibria, we introduce the following notion of pivotalness.

Definition 3. Delegate i is *pivotal* in ϕ if $Q(\phi_i', \phi_{-i}) \supset Q(\phi)$ for every $\phi_i' > \phi_i$.

That is, if a pivotal delegate ϕ_i is exchanged for a delegate $\phi_i' > \phi_i$ the set of implementable policies grows; however it need not be the case that the set shrinks if $\phi_i' < \phi_i$.[14] For example, if $N = 3$ and $\phi = (1, 1, 3)$, then $Q(\phi) = [0, 2]$. Each of ϕ_1 and ϕ_2 is pivotal: If, say, $\phi_2 = 1$ is replaced by $\phi_2' = 1.5$, then $Q((1, 1.5, 3)) = [0, 3]$. However, since only two votes are needed for a proposal to pass in ϕ, if ϕ_2 is replaced by $\phi_2' < \phi_2$ the set $Q(\phi)$ remains unchanged.

We can now state a number of results concerning the structure of representation equilibria in our model. These results characterize equilibrium legislatures and policies to some extent, and also serve as technical observations to be used later.

Lemma 2. *Let $\phi^* = (\phi_1^*, \dots, \phi_N^*)$ be any representation equilibrium. Under the assumption that $\psi_m > 0$,*

(a) No policies to the left of zero are implementable, i.e. $Q(\phi^*) = [0, 2\phi^{*(m)}]$,

(b) For all i s.t. $\psi_i \leq 0$, $\phi_i^* = \psi_i$, and for all i s.t $\psi_i > 0$, $\frac{\psi_i}{2} \leq \phi_i^* \leq \psi_i$.

(c) If $\phi_i^* \neq \psi_i$ then ϕ_i^* is pivotal in ϕ^*.

Lemma 2 (a) states that if the median voter in the median district is to the right of the status quo, then all equilibrium assemblies will be such that a change of the status quo to the left is impossible. Part (b) of the result states that the only districts that could (possibly) vote strategically and not elect their median voter as representative are those whose median voter is to the right of the status quo; this group obviously includes the median district. Further, any such strategic behavior must be conservative: If the elected representative of a district is not its median voter, then it is someone whose most preferred policy is between the status quo and the median voter's bliss point. Part (c) states that whenever a district elects a conservative representative in equilibrium, this representative must be pivotal.[15]

Call an assembly ϕ *order-preserving* if for all $i, j \in I$, $\psi_i \geq \psi_j$ implies $\phi_i^* \geq \phi_j^*$. As the term indicates, the legislators in an order-preserving assembly are (weakly) ordered in the same way as the constituencies they represent. Now define for $i \in I$ and $c \in [0, 1]$ a legislator

$$\phi_i(c) = \begin{cases} \psi_i & \text{if } i > m, \\ \min\{\psi_i, c\psi_m\} & \text{if } i \leq m, \end{cases}$$

[14] These inequalities would have to be reversed if we assumed that $\psi_m < 0$. That is, we would have to change Definition 3 to read "... for every $\phi_i' < \phi_i$."

[15] For parts (a) and (b) of Lemma 2, reverse statements would hold if $\psi_m < 0$, and for part (c) a reverse notion of pivotalness would be required.

and let $\phi(c) = (\phi_1(c), \ldots, \phi_N(c))$. $\phi(c)$ is an order-preserving assembly. Further $\phi(1) = (\psi_1, \ldots, \psi_N)$ is the "truly representative" assembly, and $\phi(c)$ with $c < 1$ is a conservative assembly. Regardless of c, all districts i such that $\psi_i < 0$ or $\psi_i > \psi_m$ will elect their median voter as representative. We then have the following result:

Theorem 2. *There exists a non-empty set $B \subset [1/2, 1]$ such that ϕ^* is an order-preserving representation equilibrium if and only if $\phi^* = \phi(c)$ for $c \in B$. Furthermore, B is closed and consists of m or fewer connected components.*

Representation equilibria exist since $B \neq \emptyset$, so there is always at least one order-preserving equilibrium. Further, order-preserving equilibria have a particularly simple structure. This structure is such that if several equilibria exist, they can be ranked by their degree of conservatism (given by the number c). The set of values c for which $\phi(c)$ is an equilibrium may have "holes," however, and the following 7-district example illustrates this property:

Example 1. Let $u(x) = -x^{5/2}$, and let $N = 7$ (so $m = 4$) and

$$\psi = (8.5,\ 9,\ 9.5,\ 10,\ 11.5,\ 12.2,\ 17.5).$$

Then, by following the construction outlined in the proof of Theorem 2, we can construct the four-component set

$$B = [.566, .572] \cup [.583, .6085] \cup [.62, .873] \cup \{1\},$$

such that ϕ^* is an order-preserving equilibrium if and only if $\phi^* = \phi(c)$ for $c \in B$. □

Note that the range of possible order-preserving equilibria in the previous example is quite large, ranging from $c = 0.566$ to $c = 1$. Consequently, the expected policy outcomes are very different across these equilibria: When $c = .566$, districts 1 through 4 elect representatives whose bliss points are at 5.66 and who will propose policy 5.66 when they are recognized, while districts 5 through 7 elect their median voters who will be constraint and propose policy 11.32. On expectation the enacted policy outcome is $E(x^*|\phi(.566)) = 8.09$. On the other hand, $\phi(1)$ is also an equilibrium outcome. This means that all districts elect their median voters and that all representatives can implement their most preferred policies when recognized. In this equilibrium the expected policy outcome is $E(x^*|\phi(1)) = 11.17$.

Order-preserving equilibria are easy to characterize, as Theorem 2 shows, and always exist. They are not the only equilibria, however. As the following example illustrates, there can be equilibria in which an order reversal occurs:

Example 2. Let $u(x) = -x$, and let $N = 7$ and

$$\psi = (-2,\ -1,\ .75,\ 1,\ 1.5,\ 3,\ 4).$$

Consider the assembly ϕ^*, given by $\phi_i^* = \psi_i\ \forall i \neq 5$, and $\phi_5^* = .75$. This assembly is not order-preserving as district 5 elects a representative to the left of district 4's representative. Nevertheless, ϕ^* is a (conservative) representation equilibrium with $Q(\phi^*) = [0, 1.5]$. To see this, consider first all districts $j \neq 5$, which elect their median voters: Districts 1 and 2 must elect their median voters by Lemma 2 (b). District 3 can make $Q(\phi^*)$ larger (but not smaller) by a unilateral change of representative, so electing $\phi_3^* = \psi_3$ is optimal. District 4 cannot change $Q(\phi^*)$ by a unilateral move, and so electing $\phi_4^* = \psi_4$ is optimal. Districts 6 and 7 can make $Q(\phi^*)$ smaller (but not larger), but since $\overline{Q}(\phi^*)$ is already to the left of ψ_6 and ψ_7 this would clearly not be in their interest. This leaves district 5, which does elect a conservative representative (who is pivotal as Lemma 2 (c) predicts) and thereby obtains utility $U_5(\phi^*) = -5$. To see that this is optimal, note that district 5 cannot make $Q(\phi^*)$ smaller, so it will not deviate to $\phi_5 < 0.75$. It can, however, enlarge $Q(\phi^*)$ up to $[0, 2]$, and the expected utility from deviating to a less conservative delegate is piecewise linear:

$$U_5(\phi^*_{-5}, \phi_5) = \begin{cases} -2.75 - 3\phi_5 & \text{if } 0.75 < \phi_5 \leq 1, \\ -6.75 + \phi_5 & \text{if } 1 < \phi_5 \leq 1.5. \end{cases}$$

It is easy to see that $U_5(\phi^*_{-5}, \phi_5) < 5 \ \forall \phi_5 \in (0.75, 1.5]$. Deviating to $\phi_5 > \psi_5 = 1.5$ can only hurt district 5, so that we have exhausted all possible deviations. □

5 Inter-District Heterogeneity and Conservatism

Conservative outcomes arise in equilibrium when a majority of constituents in moderate districts wants to constrain the representatives from more extreme districts. The examples which we presented suggest that this incentive is especially strong when the variation of median preferences across districts is high. The aim of this section is to investigate the connection between preference dispersion and conservatism in detail. To start, define

$$\chi_i \equiv \frac{\psi_i}{\psi_m} \tag{6}$$

for $i = 1, \ldots, N$. The measure χ_i represents the policy preferences in district i relative to those in the median district m as the ratio of the most preferred policies by the median voters in these constituencies. Our first result states sufficient conditions for there to be only conservative equilibria (order-preserving or otherwise), and for there to be only a unique non-conservative equilibrium:

Theorem 3. *Let $\chi_N = \psi_N / \psi_m$ measure the median preferences in the most extreme constituency relative to the median district.*

(a) If $\chi_N > 2$, every representation equilibrium exhibits conservatism.

(b) If $\chi_N < 1 + \frac{1}{N}$, the only representation equilibrium is non-conservative, i.e. $\phi^* = (\psi_1, \ldots, \psi_N)$.

As Theorem 3 shows, whether or not conservatism arises as an equilibrium phenomenon depends to a large extent on where on the political spectrum the constituencies are located *relative to each other*, rather than where they are located in absolute terms. Because it is relative rather than position that determines whether or not equilibrium legislatures are conservative, it is possible for the society to experience a preference shock that shifts all preferences in one direction, yet the distribution of implemented policies shifts in the opposite direction (in the sense of stochastic dominance):

Example 3. Let $u(x) = -x$ and $N = 5$. Compare the following configurations of constituency preferences:

(a) First, let

$$(\psi_1, \ldots, \psi_5) = (-3, -2, 4, 4.4, 4.5).$$

By Theorem 3, since $\chi_5 = \frac{4.5}{4} < 1 + \frac{1}{5}$, there is only one representation equilibrium, $\phi^* = (-3, -2, 4, 4.4, 4.5)$. The policies that arise in this equilibrium are 0 (with probability $\frac{2}{5}$), and 4, 4.4, and 4.5 (with probability $\frac{1}{5}$ each). The expected policy outcome is $E(x^*|\phi^*) = 2.58$.

(b) Next, consider

$$(\psi_1, \ldots, \psi_5) = (-2, -1, 4.2, 9, 10).$$

Now $\chi_5 = \frac{10}{4.2} > 2$, and thus every equilibrium exhibits conservatism. In particular, it can be shown that the only equilibrium is now the most conservative order-preserving assembly $\phi^* = \phi(1/2) = (-2, -1, 2.1, 9, 10)$. In this equilibrium, the policies that arise are 0 (with probability $\frac{2}{5}$), 2.1 (with probability $\frac{1}{5}$), and 4.2 (with probability $\frac{2}{5}$); a first-order stochastic shift to the left. The expected policy outcome is $E(x^*|\phi^*) = 2.1$.

The reason for the non-monotonicity in policy outcomes when going from case (a) to case (b) is that the shift in preferences to the right is accompanied by a significant increase in the heterogeneity of preferences across constituencies. The uniqueness of the equilibria in both cases makes the example particularly compelling.

The next example shows that it is also possible for expected policy outcomes to be inefficient, in the sense that it falls short of the goals of a majority of voters in every constituency.

Example 4. As in the previous example, let $u(x) = -x$ and $N = 5$. Consider the following configuration of constituency preferences:

$$(\psi_1, \ldots, \psi_5) = (3.5, 4, 4.5, 9, 10).$$

Since $\chi_5 = \frac{20}{9} > 2$, every equilibrium exhibits conservatism. However, it can be shown that now a range of conservative equilibria exists (this will be stated formally in Theorem 4 (a) below). In particular $\phi(c)$ is an equilibrium for all $\frac{1}{2} \leq c \leq \frac{8}{9}$. In the most conservative equilibrium, $\phi(1/2)$, the policies that arise are 2.25 (with probability $\frac{3}{5}$) and 4.5 (with probability $\frac{2}{5}$). The expected policy outcome is $E(x^*|\phi^*) = 3.15$. □

Thus, the expected equilibrium policy outcome can be insufficient from every constituency's perspective, at least in so far as there is a majority of voters in every district that prefers a more right-wing distribution of policies than the one that arises in equilibrium. For example, if the policy $x = 4.5$ was implemented with probability one, a majority of citizens in every constituency would be made better off relative to the policy distribution that arises in $\phi(1/2)$.[16] This alternative distribution is obviously feasible—it

[16] This majority can be arbitrarily large, for instance if almost all citizens within a constituency have ideal points at the median location. Hence, depending on the distribution of preferences within the constituencies, the conservative equilibrium can be highly inefficient, in the sense that the measure of citizens who are worse off after the policy change can be arbitrarily small.

can be achieved if every constituency elects an individual with bliss point 4.5 as their representative, but this outcome is not an equilibrium. The examples possesses many other equilibria which do note share this inefficiency. Interestingly, however, in Section 7 we will show that $\phi(1/2)$ is the only equilibrium of Example 4 which is coalition-proof.

Theorem 3 leaves room for intermediate cases, where conservative equilibria may or may not arise. To examine these situations further, we consider the special case where each citizen's utility is linear,

$$u(d) = -d,$$

and restrict our focus to order-preserving equilibria. Although linear utility implies risk neutrality in the distance between implemented policy and a voter's bliss point, the insurance motive is still present. This is the case because a voter who can shift the policy proposal of a constrained representative by a certain amount toward her own bliss point by electing a conservative representative has to "give up" only half that amount in terms of the loss incurred from the policy proposal made by her own representative. For this effect, the curvature of the utility function is obviously irrelevant; what matters is only the symmetry of preferences over policies.

Using the ratios $\chi_{m+1}, \ldots, \chi_N$, define

$$\overline{\chi} \equiv \frac{1}{m-1} \sum_{i=m+1}^{N} \chi_i. \tag{7}$$

The measure $\overline{\chi}$ represents the average dispersion of median voter preferences in districts to the right of m relative to the median voter preference in district m (recall that we assume that districts are ordered, i.e. $\psi_1 \leq \ldots \leq \psi_N$). In the linear utility case, Theorem 2 can be strengthened slightly, as B can have at most two connected components, regardless of the number of districts, stated formally in part (a) of the following result. In part (b), the theorem provides a necessary and sufficient condition for conservative equilibria to exist in part.

Theorem 4. *Assume all citizens' preferences are linear.*

(a) If the order-preserving assembly $\phi(c)$ is a representation equilibrium for some $\frac{1}{2} < c < 1$, then for all $\frac{1}{2} \leq c' < c$, $\phi(c')$ is an equilibrium.

(b) There exists an order-preserving representation equilibrium which exhibits conservatism if and only if $\overline{\chi} \geq \frac{N}{N-1}$.

The intuition for part (a) is the following. With strictly concave utility functions, the marginal cost of a conservative representative becomes larger, the higher the degree of conservatism, while the marginal benefit becomes smaller. Hence, there exists an "optimal degree of conservatism" for a legislature beyond which the additional loss of electing a slightly more conservative delegate outweighs the additional gain from doing so. With linear utility, on the other hand, these marginal costs and benefits remain unchanged, and the citizen prefers the most conservative legislature such that the set of implementable policies still contains citizen's most preferred point.

Regarding part (b) of Theorem 4, observe that $\frac{N}{N-1} \to 1$ as N becomes large. Thus, with a large number of constituencies, the inter-district heterogeneity need not be very large for conservative legislatures to be equilibrium outcomes. Keep in mind, however, that Theorem 4 does not say anything about whether the non-conservative assembly is an equilibrium or not. The following example shows that it is in fact possible, with a rather moderate degree of heterogeneity, for both $\phi(1)$ and $\phi(1/2)$ to be equilibria:

Example 5. Consider the linear utility case and suppose that $N = 21$ districts are uniformly distributed on $[4, 5]$, i.e.

$$\psi_1 = 4,\ \psi_2 = 4.05,\ \ldots,\ \psi_m = \psi_{11} = 4.5,\ \ldots,\ \psi_{20} = 4.95,\ \psi_{21} = 5.$$

There exists a non-conservative equilibrium in which every constituency elects its median voter. However,

$$\overline{\chi} = 1.05\overline{5} > 1.05 = \frac{N}{N-1}.$$

Thus there is also a representation equilibrium $\phi^* = \phi(1/2)$ in which $m = 11$ districts elect a representative with bliss point 2.25. □

As in Example 4, a majority of voters in each district agrees that a change of the status quo policy toward the right is desirable. Unlike in the previous example however, there is now considerable less disagreement across district median voters about the ideal; in particular $\frac{8}{9} \leq \chi_i \leq \frac{10}{9}$ for all i. Despite the fact that there is almost a consensus among the district median voters regarding the optimal policy outcome, the remaining disagreement is large enough for conservative equilibria to exist in addition to the non-conservative equilibrium. The possible impact of conservatism on enacted policy is not negligible: In $\phi(1)$ the expected policy outcome is $E(x^*|\phi(1)) = 4.5$, while in $\phi(1/2)$ it is $E(x^*|\phi(1/2) = 3.375$. It is, however, not the case anymore that only the most conservative equilibrium is coalition-proof: As shown in Section 7, $\phi(1/2)$ is not immune to coordinated deviations across districts, while $\phi(1)$ is.

6 Empirical Observations

Our model and its results illuminate a set of political phenomena from a new perspective, which we discuss in this section. In particular, we argue that our model can provide possible explanations for the following observations: The apparent unrepresentativeness of the U.S. congress, the sluggishness with which economic reforms are being implemented in many European countries despite the fact that there exists a rather strong consensus within the citizenry that reforms are needed, and so-called backlash effects that were the result of increased minority representation in several U.S. state legislatures. These are rather disparate observations, but all of them concern distortions of outcomes, such as actual enacted policy or actual political representation, relative to the distribution of underlying characteristics of the electorate. Of course, many explanations exist for each of these effects, and we regard our results as complementary to the existing explanations. In particular, it is our view that all three phenomena are larger puzzles with many contributing factors, and the argument put forth in this section is that taking a closer look at constituency heterogeneity in a framework of indirect legislative policy making can provide valuable new insights into each puzzle.

6.1 Unrepresentative representatives

As we have argued, our model is a very stylized account of real legislative elections and policy making. Is there any evidence that the type of strategic reasoning that leads to conservatism in our model is employed by voters in the real world? To address this question head-on, it seems that one ought to measure policy preferences of voters in constituencies such as U.S. congressional districts, and compare them with the preferences of their elected representatives. Miller and Stokes (1966) compiled the first, and still widely used, set of data concerning the preferences of U.S. citizens and their representatives. Their seminal study found little correlation between attitudes of representatives of U.S. Congress of 1958 and the *mean* (not median!) attitude of their constituents on the issues of civil rights legislation and social welfare. These findings may have been be exaggerated, however, by the fact that only very few constituents of a congressional district were interviewed and thus a large amount of noise is present in the survey data. Erikson (1978) reexamined Miller and Stokes' original paper, accounting for noise and possible measurement errors, and found a slightly higher correlation between representatives and constituent attitudes. Bartels (2002) uses data on constituents' opinions from the 1988–1992 Senate Election Study and voting behavior of senators in the 101st–103rd U.S. Congress (1989–1994). He finds that senators' voting behavior is mostly explained by the attitudes of their rich constituents. This pattern can be for a variety of reasons, including non-participation of poor constituents in congressional elections, the fact that wealthy citizens are more likely to be personally acquainted with their senators, campaign contributions made by wealthy constituents to their senators, or attempts to please wealthy swing voters who are most likely to change their vote in response to changes in economic policy platforms.

Our model suggests a further explanation for the empirical findings, namely that Congress is unrepresentative because of a conservatism effect caused by a high degree of heterogeneity across constituencies. Consider economic policy. In many cases, voter preferences over economic policy depend on underlying economic characteristics such as their wealth or income, and these characteristics vary greatly across constituencies. For example, the median household income across U.S. congressional districts in 1999 ranged from $19,311 (New York district 16, *Bronx*) to $80,391 (Virginia district 11, *Fairfax*).[17] The median household income in the richest district is thus more than four times as high as the corresponding figure in the poorest district, and almost twice as high as the median household income in the median district ($41,060). The median household income across U.S. States naturally exhibits less dispersion, but it still ranges from $29,696 in West Virginia to almost twice as much, $55,146 in New Jersey (again for the year 1999). Hence it is likely that the median voters in poor constituencies have preferences over economic policy which are very different from those of median voters in rich constituencies. Constituencies with a large fraction of poor voters benefit the most from federal programs such as social security, and a majority of their citizens may prefer an increase in the size of these programs. The opposite is true rich voters, who mostly pay for such programs. In constituencies of moderate median income or wealth levels, rich legislators may therefore be strategically elected to limit the risk of excessive federal spending on welfare programs

[17] In 1999 dollars; Source: U.S. Census Bureau.

or social security.[18]

6.2 Reform obstacles

In our model, enacted policy is simply a change of the status quo. A natural interpretation of such a change is that it represents a policy reform. Example 3 shows that it is possible for every citizen's view to shift to the right, but for policy outcomes to shift to the left on expectation (and vice versa). This happens if the shift in political preferences is accompanied by a simultaneous increase in preference dispersion that is large enough for conservative outcomes to prevail after the shift. In Example 4, it is further shown that economic reforms may fall short of the policy goals of a majority of citizens. Thus, even when many citizens stand to gain from a reform, if the disagreement regarding the optimal magnitude of the reform is large the eventual outcome may be too little reform. We view these effects as a possible explanation for why some countries have been experiencing great difficulties in implementing economic and social reforms that most citizens agree are needed. Proposals to reform trade policy or social security systems, for example, are often met with substantial resistance even though the need for such reforms, and the benefits associated with it, lay abundantly clear.

Consider the case of trade liberalization, and suppose a country needs to decide on the degree of protection for domestic producers. Suppose further that in every electoral district the median voter's preference shifts in the direction of fewer restrictions on trade. This may be the case, for example, when the benefits from foreign trade become apparent after an initial round of tariff reductions. In districts where a large share of the population is employed in industries most threatened by further liberalization, however, the shift may be modest compared to districts that stand to gain the most from it, and it is possible that the final outcome is, in expectation, a rollback of the initial tariff reductions. It is also possible for there to be broad popular support for significant trade liberalization, and while such liberalization occurs in equilibrium it falls short of the policy goals of most citizens. Another current case of reform obstacles is provided by some large European economies, most notably France, Germany, and Italy. These economies have been struggling greatly over the last several years to modernize their costly welfare states, while there has been seemingly little disagreement among their citizens that reform is needed in order to sustain these welfare systems in the long term. In Germany, for example, a rather broad consensus has existed for many years that a reform of rigid labor laws is desirable, and while some modest reforms have been implemented by the government they were promptly criticized by many economists as insufficient to reduce Germany's high unemployment rate.

The idea that an uneven distribution of the benefits from reform projects presents an obstacle to reform is not new.[19] It is well-known that the outcomes of democratic processes may not be optimal in the aggregate if individuals are heterogeneous in certain characteristics, such as the types of factors they own (Alesina and Rodrik, 1994). Even if a majority stands to gain from a policy reform, strong political organization of the losing individuals through special interest groups can effectively prevent reform (Brock and Magee, 1978), and simple majority voting in a direct democracy cannot implement

[18]In this case, a conservative legislature is also conservative in the "small government" sense.

[19]See Rodrik (1996) for a detailed a survey of the literature on economic reform.

the reform unless the losing majority is compensated. Fernandez and Rodrik (1991) show that when there is uncertainty about the identity of those who will benefit and those who will lose from a reform, there can be majoritarian outcomes in which the reform project is not undertaken, while it is common knowledge that ex-post (when these identities are revealed) it would have majority support. Messner and Polborn (2004) examine a dynamic model with a time lag between the accrual of costs and benefits of reforms, and derive a rationale for a constitution that requires super-majorities in order to implement reforms. In their model, too, it is generally not sufficient that the median voter prefers a reform project over the status quo for the reform to be implemented.

The mechanism we propose, by which this unevenness of reform benefits translates into legislative reform obstacles, is new and adds to the reform resistance literature. Through their votes in legislative elections, citizens not only determine the policies advocated on their behalf, but can also shape the post-election constraint set for the legislature. As we have shown, this possibility acts as insurance against too much reform, and can be more important in an individual's cost-benefit calculus than the direct advocacy effect of one's representative. It is instructive to compare our approach to the Fernandez and Rodrik (1991) model of direct democracy. Like theirs, our model relies on uncertainty to make reforms more difficult to implement. However, the uncertainty in our model stems from the complexities of the legislative process: Citizens know how much they will lose or benefit from a particular reform, but do not know exactly what the proposed reform will be. In Fernandez and Rodrik (1991) there is a single reform project up for vote whose true effects are unknown to the citizens, which can lead to an impasse in implementing the reform. In the current paper, this impasse is less complete: As long as a majority of citizens in a majority of constituencies prefers some reform in the same direction, there will be some reform on expectation—however, in a conservative equilibrium the expected amount of reform will be less than what it would be if each constituency was represented by its median individual. It may then be possible to make a majority of citizens in every constituency better off by implementing more reform, and one can argue that such an incomplete reform deadlock reflects the current situation in the aforementioned European economies.

6.3 Minority representation and backlash effects

The model also provides a formal underpinning of the so-called "backlash theory" in sociology. According to this theory, increased diversity within organizations may lead to a decrease in the likelihood of members of the dominant group (whites, men, etc.) adopting behavior that advances the interests of the disadvantaged group (blacks, women, etc.). There is empirical evidence that such a backlash has indeed happened in several state legislatures in the U.S. as representation of female and African-American voters has increased over the last three decades. Bratton (2002) examines the composition of six U.S. state legislatures over three decades with respect to the fraction of women and minority members, and uses the diversity of these bodies to explain agenda-setting behavior of legislators. She finds that as legislative bodies have become more diverse, the number of bills introduced to advance minority interests has increased, but the number of such bills introduced by majority legislators has decreased. Our model provides a rational explanation of backlash effects in legislatures: An increase in the share of minority members, who are likely to possess very different policy goals than most members, leads to an increased risk of extreme legislation being passed. As a result, some constituents may find it optimal to vote for representatives who are less inclined to vote for a particular bill that advances minority interests, and thus are also less likely to introduce such bills themselves. Note that for such an effect to take place it is not necessary that the fraction of minority individuals among the population in an electoral district grows, but only that more minority individuals vote in legislative elections. For instance, as the turnout of African-American voters in southern states has grown since the civil rights era, the fraction of black legislators in many state assemblies has increased as well. Predominantly white electoral districts who previously elected moderates may have responded by electing more conservative white representatives to constrain black legislators in their agenda-setting activities. The result is then a racially diverse legislature, with both an increase in the number of black-interest bills introduced by black legislators, and a decrease in black-interest bills sponsored by whites.[20]

7 Coalition-Proof Representation Equilibria

Theorem 2 identifies a set of equilibria which may contain more than a single element. Examples 1, 3, 4, and 5, confirm that the case of multiple equilibria is indeed quite common in the model: In each of the examples there exists a continuum of representation equilibria that can be ranked by their degree of conservatism. This multiplicity of equilibria arises since conservative equilibrium legislatures tend to be highly coordinated outcomes: By part (c) of Lemma 2, all representatives who are not of the same type as their median constituent share the same preferences. There are many such coordinated profiles in which changing a single district's representative does not increase the utility of a majority of citizens in this district; hence these profiles are all equilibria.

To deal with this multiplicity, we now apply the refinement concept of *coalition-proof Nash equilibrium (CPN)*, introduced by Bernheim, Peleg, and Whinston (1987). Roughly, an equilibrium is coalition-proof if it is not upset by joint deviations of coalitions of players.[21] Implicit in the assumption that multilateral deviations are possible is the idea that players can coordinate their actions not only in equilibrium, but also out of equilibrium. However, if this assumption is made one should also allow for the possibility that subgroups of deviating coalitions further coordinate their actions and "deviate from the deviation." Hence, in order to upset an outcome a coalitional deviation must be credible, in the sense that the deviation itself is immune to further coalitional deviations from it. (without this second requirement, we would have *strong equilibria*.) The formal definition of a credible deviation, and thus of coalition-proofness, is recursive: Let $\phi \in \mathbb{R}^N$ be a strategy profile and let $C \subseteq I$ be a coalition of players. Denote by $\phi'_C \subseteq \mathbb{R}^C$ a coalitional deviation for C. Then ϕ'_C is a credible deviation from ϕ for C if the following holds: If $|C| = 1$, ϕ'_C is

[20] Bratton performs separate analyses for Democratic and Republican legislators, whereas for the current purpose a pooled analysis including all legislators of an assembly seems more appropriate. Nevertheless, the findings are indicative of increased diversity having resulted in a backlash effect.

[21] A coalition-proof equilibrium is automatically a Nash equilibrium, since the set of possible deviations against which the outcome must be checked contains unilateral deviations.

credible if it improves the payoff for the single player in C strictly. If $|C| > 1$ then ϕ'_C is credible if every member of C obtains a weakly higher payoff from (ϕ'_C, ϕ_{-C}) than from ϕ, some member obtains a strictly higher payoff, and there does not exist $C' \subseteq C$ and $\phi''_{C'} \in \mathbb{R}^{C'}$ such that $\phi''_{C'}$ is a credible deviation from (ϕ'_C, ϕ_{-C}) for C'. An outcome ϕ is then a coalition-proof equilibrium if no coalition $C \subseteq I$ has a credible deviation.

In the context of our model, the coalition-proof criterion not only serves as a selection device when there are multiple equilibria, but has practical significance. Recall that in any representation equilibrium, each district's representative maximizes the district median voter's utility conditional on the set of representatives elected in the other districts. Since such coordination is not a trivial task, the question arises what mechanism would allow for coordination of voters across districts to be successful. A natural institution that can resolve this coordination problem is that of a well-functioning political party system. Firstly, the nomination process within a party can ensure that in each district expected to elect a conservative representative, such a candidate is available to the voters, and that the candidates across districts are very similar in their political stance (otherwise some of them would not be pivotal if elected). Secondly, the party can utilize its resources to make sure that these candidates are visible to the voters *across* districts. We do not model the nomination process in this paper, nor do we model political parties. However, we strongly suspect that if a particular outcome is not a coalition-proof equilibrium, it is unlikely to emerge in a world where those institutions exist.[22]

We now state a necessary condition for a representation equilibria to be coalition-proof, and identify one such equilibrium. Let $B \subseteq [1/2, 1]$ be the set of values c such that $\phi(c)$ is an order-preserving equilibrium if and only if $c \in B$ (i.e. the set identified in Theorem 2). Let $B^* = \arg\max_{c \in B} U_m(\phi(c))$ be the set of equilibria that maximize the utility of the median voter in district m within the set of representation equilibria.

Theorem 5. *Suppose B^* is a singleton, $B^* = \{c^*\}$.*

(a) $\phi(c)$ is a coalition-proof order-preserving equilibrium only if $c \leq c^*$.

(b) $\phi(c^*)$ is a coalition-proof equilibrium.

Thus, an order-preserving assembly cannot be a coalition-proof outcome if it lies to the right m's most-preferred order-preserving equilibrium. This condition is stated in terms of the degree of conservatism of an order-preserving legislature, so an outcome which is rejected based on Theorem 5 (a) is not conservative enough. Theorem 5 does not fully characterize the set of coalition-proof equilibria, since it provides no condition that is necessary and sufficient at the same time, and because it does not apply to the case where B^* is not a singleton.[23] However, the necessary condition provided in part (a) is easily checked and relatively strong; for instance it is strong enough to rule out all but one equilibrium in Example 4 (see below). Furthermore, since $U_m(\phi(c))$ is continuous in c and

[22] In the model of Morelli (2004), the role of parties is similar: They coordinate the individual candidates' campaign platforms as well as voters' actions.

[23] If u is strictly concave, the case that B^* contains multiple elements is non-generic, even though B can be a continuum. However, in the linear case $u(d) = -d$, U_m can easily have flat parts, in which case it is generically possible that B^* is a continuum.

B is compact by Theorem 2, $B^* \neq \emptyset$. Thus, as long as B^* contains only one element, part (b) implies that a coalition-proof equilibrium exists.

Theorem 5 can be directly applied to our previous examples. In Example 4, the most conservative equilibrium is the unique equilibrium in which the utility of the median voter in district m is maximized. Thus it must be the unique coalition-proof outcome, and the inefficiency that arises in this equilibrium cannot easily be resolved by "re-coordination" through institutions such as political parties. On the other hand, none of the conservative equilibria in 5 are coalition-proof. To see this, note that $\phi(1)$ is the most-preferred equilibrium of the median voter in district $m = 11$, so by Theorem 5 (b) it is coalition-proof. Since $\phi(1)$ is also the most preferred equilibrium of all districts $1 \leq j \leq 10$, in $\phi(c)$, $c < 1$, the coalition $D = \{1, \ldots, 11\}$ can improve each member's payoff by agreeing to elect their median voters ψ_j as representative, instead of the conservative representative $c\psi_m$. This will result in assembly $\phi(1)$ after the deviation, and since this outcome is coalition-proof the initial deviation from $\phi(c)$ is credible.

8 Discussion

We presented a model of indirect representative democracy that incorporates the dual role of elected representatives: To advocate policy, and to enact policy. We were able to characterize some properties of representation equilibria in general, and to identify a class of simple, order-preserving equilibria which always exist and each of which is characterized by a single number $c \in B \subset [1/2, 1]$. The actual set of equilibria will depend on the degree of inter-district heterogeneity. As we have shown, if it is large enough conservative equilibria exist and a non-conservative equilibrium may no longer exist.

The paper has a number of promising extensions, not pursued in this paper. For example, the model can be made dynamic by repeating it indefinitely, with the enacted policy in one round becoming the status quo in the next round. If citizens are not forward-looking, it seems likely that the process eventually reaches a "steady state," with the the median ideal point in he median constituency as the unique policy outcome. This process may be very slow, however, and examining its speed would provide a framework for studying issues of "slow reform," or piecemeal approaches to reform. Another extension is to eschew the exogenously fixed constituencies and replace them with a single, multi-member district. This would necessitate the introduction of a suitable equilibrium concept for multi-candidate elections, to replace the Condorcet-type pairwise comparisons we use to define equilibrium in this paper. The concept of a "constituency" would then become endogenous. This approach may better describe European democracies than the current one, and perhaps lead to interesting comparative insights.

Appendix: Proofs

To simplify some lengthy expressions, throughout the appendix we multiply all expected utilities by N so that the factor $1/N$ can be cancelled from the utility terms.

Proof of Lemma 1

Note that $U_\psi(\phi)$ is defined in (5) as the sum of functions of the form $u(|z_i(\phi) - \psi|)$, where z_i is weakly increasing in ϕ as shown in the main text, and u is concave. Because z_i is not concave, $U_\psi(\phi)$ is not necessarily concave in ϕ; however $U_\psi(\phi)$ is concave in ψ given ϕ. The monotonicity of z_i together with the concavity of U_ψ in ψ then implies that U_ψ exhibits dereasing differences: For $\phi \geq \phi'$ and $\psi \geq \psi'$, we have

$$U_\psi(\phi) - U_\psi(\phi') \geq U_{\psi'}(\phi) - U_{\psi'}(\phi'). \tag{8}$$

Since $z_i(\phi) \in Q(\phi)$ by definition and $Q(\phi)$ is closed, it follows that $U_\psi(\phi) \leq U_\psi(\phi')$ for $\psi < \underline{Q}(\phi')$ and $U_\psi(\phi) \geq U_\psi(\phi')$ for $\psi > \overline{Q}(\phi)$. By (8), then, there must exist a value $\hat{\psi} \in [\underline{Q}(\phi'), \overline{Q}(\phi)]$ such that $U_\psi(\phi) \leq U_\psi(\phi')\ \forall \psi \leq \hat{\psi}$, and $U_\psi(\phi) \geq U_\psi(\phi')\ \forall \psi \geq \hat{\psi}$. $\square$

Proof of Lemma 2

We proceed in five steps that build on each other. Assertion (a) of the result then follows from step 1, (b) follows from steps 2, 3, and 5 together, and (c) follows from step 4.

Step 1: $\underline{Q}(\phi^*) = 0$ and thus $Q(\phi^*) = [0, 2\phi^{*(m)}]$

Suppose instead that $Q(\phi^*) = [\underline{Q}(\phi^*), 0]$ with $\underline{Q}(\phi^*) < 0$. This means that at least m elements in ϕ^* are to the left of zero, but at the same time at least m of the district meadian voters have bliss points weakly to the right of zero. Hence there must be at least one district, say i, such that $\psi_i \geq 0$ but $\phi_i^* < 0$. Note that when i's delegate ϕ_i^* has agenda power, $\underline{Q}(\phi^*) < 0$ implies that i's policy proposal satisfies $z_i(\phi^*) < 0$. Now suppose i elected $\phi_i' = 0$ instead of $\phi_i^* < 0$, and denote by ϕ' the new assembly, i.e. $\phi' = (\phi_{-i}^*, \phi_i')$. Then $Q(\phi') \subseteq Q(\phi^*)$, which in turn means that all policy proposals that are made to the assembly ϕ' are weakly closer to ψ_i those under ϕ^*. Furthermore, if i's delegate has agenda power the policy proposal $z_i(\phi') = 0$ is strictly closer to ψ_i than the proposal $z_i(\phi^*) < 0$. Thus $U_{\psi_i}(\phi') > U_{\psi_i}(\phi^*)$, so that $\phi_i^* < 0$ is defeated in the election in district i by $\phi_i' = 0$, and ϕ^* cannot be an equilibrium. Thus all i with $\psi_i \geq 0$ will elect delegates $\phi_i^* \geq 0$, and since there are at least m of these districts, $Q(\phi^*) = [0, 2\phi^{*(m)}]$.

Step 2: For all $\psi_i < 0$, $\phi_i^* = \psi_i$

Consider any i with $\psi_i < 0$, and let $\phi' = (\phi_{-i}^*, \phi_i')$ and $\phi'' = (\phi_{-i}^*, \phi_i'')$. Clearly $U_{\psi_i}(\phi') > U_{\psi_i}(\phi'')$ if $\phi_i' = 0 < \phi_i''$; so it must be that $\phi_i^* \leq 0$. Since $Q(\phi^*)$ does not contain points to the left of zero (by Property 1), the median voter in district i (with $\psi_i < 0$) is indifferent between all $\phi_i^* \leq 0$, and so by Assumption 1 he elects $\phi_i^* = \psi_i$.

Step 3: For all $\psi_i \geq 0$, $\phi_i^* \leq \psi_i$

Consider any i with $\psi_i \geq 0$. We first show that $\phi_i^* \leq \psi_i$. Suppose the contrary that $\phi_i^* > \psi_i$, and consider a switch to $\phi_i' = \psi_i$; call the resulting assembly ϕ'. If $\psi_i \geq \overline{Q}(\phi^*) = 2\phi^{*(m)}$, then $2\phi^{*(m)} = 2\phi'^{(m)}$ and thus $Q(\phi^*) = Q(\phi')$; neither ϕ_i^* nor ψ_i are in this set. The distribution over policy outcomes is therefore not affected by the switch and neither is the utility of the median voter in district i. By invoking Assumption 1 we conclude that $\phi_i' = \psi_i$ should have been elected instead of $\phi_i^* > \psi_i$. On the other hand, if $\psi_i < \overline{Q}(\phi^*)$ it is possible that $Q(\phi') \subset Q(\phi^*)$. Then, however, it must also be true that $\psi_i < \overline{Q}(\phi')$. Otherwise, since $\overline{Q}(\phi') = 2\phi'^{(m)}$ we would have $\psi_i \geq 2\phi'^{(m)}$, but since $\phi_i^* > \phi_i'$ this just means that $2\phi^{*(m)} = 2\phi'^{(m)}$. Hence $\psi_i \geq 2\phi^{*(m)} = \overline{Q}(\phi^*)$, a contradiction, and therefore $\overline{Q}(\phi') > \psi_i$. But then $U_{\psi_i}(\phi') > U_{\psi_i}(\phi^*)$, since under assembly ϕ' all policy proposals $z_i(\phi')$ are weakly closer to ψ_i, and strictly closer in at least one case, namely when i is agenda setter. We again conclude that $\phi_i' = \psi_i$ should have been elected instead of $\phi_i^* > \psi_i$. Therefore, $\phi_i^* \leq \psi_i$ for all i such that $\psi_i \geq 0$. For the special case that $\psi_i = 0$, this means that $\phi_i^* = 0$.

Step 4: If $\phi_i^* \neq \psi_i^*$, then ϕ_i^* is pivotal in ϕ^*

From steps 2 and 3, the only possibility that $\phi_i^* \neq \psi_i$ is that $\psi_i > 0$ and $0 \leq \phi_i^* < \psi_i < \overline{Q}(\phi^*)$. Suppose this is the case for some i, but ϕ_i^* is not pivotal in ϕ^*. Consider a switch to $\phi_i' = \phi_i^* + \varepsilon$ and let $\phi' = (\phi_{-i}^*, \phi_i')$ be the new assembly. By the definition of pivotalness, there must exist $\varepsilon > 0$ such that $Q(\phi') = Q(\phi^*)$ and $\phi_i' \leq \psi_i$. But then arguing as before, we have $U_{\psi_i}(\phi') > U_{\psi_i}(\phi^*)$, and ϕ_i' should have been elected instead of ϕ_i^*. Thus if $\phi_i^* \neq \psi_i$ then ϕ_i^* is pivotal in ϕ^*.

Step 5: For all $\psi_i > 0$, $\phi_i^* \geq \psi_i/2$

Suppose $\psi_i > 0$ and $\phi_i^* < \psi_i/2$; then by step 4 ϕ_i^* is pivotal in ϕ^*. But using step 1, this implies that $\overline{Q}(\phi^*) < \psi_i$. Consider a switch to $\phi_i' = \phi_i^* + \varepsilon$ and let $\phi' = (\phi_{-i}^*, \phi_i')$ be the new assembly. Then, by pivotalness and the fact that $\phi^{(m)}$ is continuous in ϕ, there must exist $\varepsilon > 0$ such that $\overline{Q}(\phi^*) < \overline{Q}(\phi') = 2\phi'^{(m)} \leq \psi_i$. But then arguing as before, we have $U_{\psi_i}(\phi') > U_{\psi_i}(\phi^*)$ so ϕ^* is not an equilibrium. □

Proof of Theorem 2

The proof is in a sequence of steps which we will outline here. The detailed steps are presented below. Let ϕ^* be an order-preserving representation equilibrium. First we show that $\phi^* = \phi(c)$ for some $c \in [1/2, 1]$; hence an order-preserving equilibrium can be characterized by a single number c. This is done in Step 1. Next we characterize the set B of equilibrium values c. To this end we partition $[1/2, 1]$ into m or fewer subintervals (the precise number depends on the configuration of the ψ_i's), and inspect the expected utility for agents from the assembly $\phi(c)$ for c within each of the so constructed intervals. This is done in Steps 2 and 3. We then examine possible deviations from $\phi(c)$ and show that within each of these subintervals, the set of values c which are equilibria is a closed interval (including the possibility of an empy set empty or a singleton set). This is done in Steps 4–6 (in Step 4 we look at those cases where at least two representatives are pivotal, in

Step 5 at those cases where there is only one pivotal representative, and Step 6 examines a special case where both may happen). Thus B is as described in the statement, and ϕ^* is an order-preserving equilibrium if and only if $\phi^* = \phi(c)$ for $c \in B$. Finally, in Step 7 we argue that B is non-empty.

We will adopt the following notation throughout the proof: Since all equilibria which we consider are of the form $\phi(c)$, when no confusion arises we say "c is an equilibrium" instead of "$\phi(c)$ is an equilibrium." Likewise if c is a candidate equilibrium and we consider a deviation by a single district, say m, from $\phi_m(c) = c\psi_m$ to $\phi_m' = c'\psi_m$, we may simply say "m deviates to c' " instead "m deviates to $\phi_m' = c'\psi_m$." Given $c, c' \in [1/2, 1]$, we let $\phi(c, c') \equiv (\phi_{-m}(c), \phi_m(c'))$; this is the assembly that results when m deviates to c' in the profile $\phi(c)$. If $c' > c$ we speak of an upward deviation, and if $c' < c$ of a downward deviation.

Step 1: ϕ^* is an order-preserving equilibrium implies $\phi^* = \phi(c)$ for some $c \in [1/2, 1]$

Notice that by Lemma 2 (b) we have $\phi_i^* = \psi_i$ if $\psi_i < 0$ in any equilibrium. Furthermore, if we focus on order-preserving equilibria then $\phi^{*(m)} = \phi_m^*$ and thus $\overline{Q}(\phi^*) = 2\phi_m^*$. This implies that $\phi_i^* = \psi_i$ for all $\psi_i > \psi_m$: Suppose instead that $\phi_i^* < \psi_i$ for some $\psi_i > \psi_m$. Then by Lema 2 (c) ϕ_i^* must be pivotal, but this requires that $\phi_i^* < \phi_m$ which is impossible if ϕ^* is order-preserving. So if $\phi_i^* < \psi_i$ it must be that $0 < \psi_i \leq \psi_m$, and by Lemma 2 (b) all $\phi_i^* < \psi_i$ must be pivotal. This implies $2\phi_i^* = \overline{Q}(\phi^*) = 2\phi_m^*$, or $\phi_i^* = \phi_m^*$. Hence $\phi^* = \phi(c)$ for some $c \in [0, 1]$. If $c < 1/2$ then $\phi_m(c) < \psi_m/2$, contradicting Lemma 2 (b). We therefore conclude that $c \in [1/2, 1]$.

Step 2: The partition of $[1/2, 1]$ into subintervals

Let $\chi_i \equiv \psi_i/\psi_m$, and let $K = \max\{i \in I : \chi_i < 2\}$. We will consider the following closed intervals in which c can be contained:

$$C_m \equiv \left[\frac{1}{2}, \frac{1}{2}\chi_{m+1}\right], \; C_{m+1} \equiv \left[\frac{1}{2}\chi_{m+1}, \frac{1}{2}\chi_{m+2}\right], \ldots, C_K \equiv \left[\frac{1}{2}\chi_K, 1\right].$$

(Technically, since these intervals overlap at the end points, the collection $\{C_m, \ldots, C_K\}$ is not a partition, but the terminology is immaterial to the argument given below. It is important, however, that each of the intervals is closed.) The interpretation of these intervals is the following: If $c_1, c_2 \in \text{int}(C_k)$, then the set of representatives that are constrained in $\phi(c_1)$ resp. $\phi(c_2)$ (in the sense that their bliss points cannot be implemented) are identical, namely these representatives are those from districts $k+1, \ldots, N$. Note that there can be at most m such intervals, and $\cup_k C_k = [1/2, 1]$. Figure 3 illustrates the construction of the intervals $C_K, \ldots, C_m$.

For any $c \in [1/2, 1]$, let $d(c) = \{i \in I : \phi_i(c) \leq \psi_i\}$ be the set of districts with pivotal delegates; $d(c)$ will be of the form $\{l, \ldots, m\}$ for some $l \leq m$. (For example in Figure 3, take $c \in C_m$, then $d(c) = \{m-2, m-1, m\}$.) Each $i \in d(c)$ can enlarge the constraint set $Q(\phi(c))$ by electing $\phi_i' > c\psi_m$. We distinguish two cases, $|d(c)| = 1$ and $|d(c)| > 1$, and $c = \chi_{m-1}$ is the cutoff-point above which $|d(c)| = 1$ and at or below which $|d(c)| > 1$.

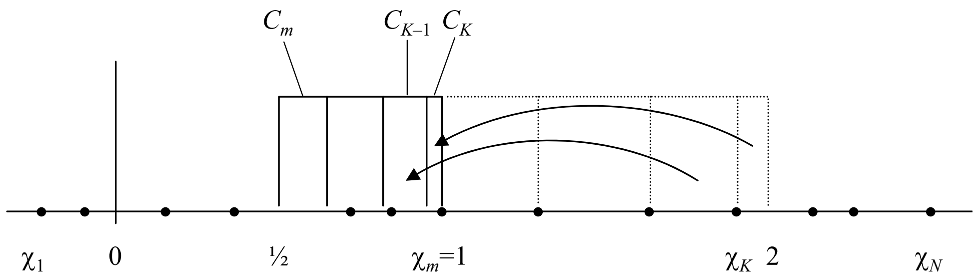

Figure 3: Construction of the intervals $C_m, \ldots, C_K$

Now define

$$\kappa \equiv \begin{cases} m-1 & \text{if } \chi_{m-1} < 1/2, \\ k \in \{m, \ldots, K-1\} & \text{if } \chi_k/2 \leq \chi_{m-1} < \chi_{k+1}/2, \\ K & \text{if } \chi_K/2 \leq \chi_{m-1} \leq 1. \end{cases}$$

The value κ represents the interval in which this cutoff point lies. For example in Figure 3, $\chi_{m-1} \in C_{K-1}$, so $\kappa = K-1$. Specifically, κ will be such that $\forall c \in C_k$, $|d(c)| = 1$ if $k > \kappa$ and $|d(c)| > 1$ if $k < \kappa$. For $c \in C_\kappa$, $|d(c)| > 1$ iff $c < \chi_{m-1}$.

Step 3: Inspection of $U_m(\phi(c))$ for $c \in C_k$

For each of the intervals defined so far, m's expected utility from assembly $\phi(c)$ can then be written as

$$U_m(\phi(c)) = \sum_{i>k} u((2c-1)\psi_m) + \sum_{m<i\leq k} u(\psi_i - \psi_m)$$
$$+ \sum_{i \in d(c)} u((1-c)\psi_m) + \sum_{i:0<\psi_i<c\psi_m} u(\psi_m - \psi_i) + \sum_{i:\psi_i \leq 0} u(\psi_m). \quad (9)$$

Observe that (9) consists of five terms. The first summation term represents m's utility if an agenda setter is selected whose bliss point is to the right of $Q(\phi(c))$. Similarly, the second term applies if an agenda setter is selected who is to the right of ψ_m but can implement his most preferred policy, the third term if a delegate in $d(c)$ (including m) is the agenda setter, the forth term if an agenda setter is selected who is to the left of ψ_m and can implement his most preferred policy, and the last term if an agenda setter is selected who prefers a policy to the left of the status quo (which means the status quo is maintained).

Note that (5) is continuous in c, but will have kinks at those c where $|d(c)|$ has jumps. Where possible we differentiate (9) twice with respect to $c \in C_k$ to obtain

$$\frac{\partial}{\partial c} U_m(\phi(c)) = 2\psi_m \sum_{i>k} u'((2c-1)\psi_m) - \psi_m |d(c)| u'((1-c)\psi_m), \quad (10)$$

$$\frac{\partial^2}{\partial c^2} U_m(\phi(c)) = 4\psi_m^2 \sum_{i>k} u''((2c-1)\psi_m) + \psi_m^2 |d(c)| u''((1-c)\psi_m) \leq 0. \quad (11)$$

Since $|d(c)|$ decreases in c, m's utility from profile $\phi(c)$ is concave within each interval C_k.

Step 4: Equilibria in C_k ($k < \kappa$)

Pick any interval $C_k = [\chi_k/2, \chi_{k+1}/2]$ with $k < \kappa$. Define

$$C_k^* = \{c \in C_k : \phi(c) \text{ is an order-preserving representation equilibrium }\}.$$

We will now characterize C_k^*. Lemma 1 implies that if m has no incentive to deviate upwards, then no $i \in d(c)$ has. Hence, to check if $c \in C_k^*$ we only need to verify that median voter ψ_m does not prefer to deviate to $c' > c$. Since $|d(c)| > 1$ we do not have to consider downward deviations to $c' < c$: Such a representative would not be pivotal, so the only effect such a deviation has on m's expected utility is that in the case m is agenda setter, a policy farther away from ψ_m will be implemented, which reduces m's expected utility. In addition, as far as upward deviations are concerned we do not have to consider $c' > 1$, since if such a deviation was profitable for m then $c' = 1$ would be profitable as well. Hence we only consider deviations to $c' \in (c, 1]$. We will break the analysis up into two substeps.

Step 4a. Suppose first m deviates upward within C_k, i.e. to $c' \in C_k$, $c' > c$. Note that $Q(\phi(c)) \subset Q(\phi')$ since $\phi_m(c)$ is pivotal in $\phi(c)$. The expected utility for m after the deviation can be written as

$$\begin{aligned} U_m(\phi(c, c')) = \sum_{i>k} u((2c'-1)\psi_m) + \sum_{m<i\leq k} u(\psi_i - \psi_m) + u((1-c')\psi_m) \\ + \sum_{i \in d(c)-m} u((1-c)\psi_m) + \sum_{i:0<\psi_i<c\psi_m} u(\psi_m - \psi_i) + \sum_{i:\psi_i \leq 0} u(-\psi_m). \quad (12) \end{aligned}$$

(12) is continuous and differentiable in c', and differentiating twice we obtain the following derivatives:

$$\frac{\partial}{\partial c'} U_m(\phi(c, c')) = 2\psi_m \sum_{i>k} u'((2c'-1)\psi_m) - \psi_m u'((1-c')\psi_m) \quad (13)$$

$$\frac{\partial^2}{\partial c'^2} U_m(\phi(c, c')) = 4\psi_m^2 \sum_{i>k} u''((2c'-1)\psi_m) + \psi_m^2 u''((1-c')\psi_m) \leq 0. \quad (14)$$

Observe that both derivatives are independent of c, and $U_m(\phi(c, c')$ is concave in c'. These facts imply that if

$$\left.\frac{\partial}{\partial c'} U_m(\phi(c, c'))\right|_{c'=c} \leq 0 \;\;\Rightarrow\;\; \left.\frac{\partial}{\partial c'} U_m(\phi(\tilde{c}, c'))\right|_{c'=\tilde{c}} \leq 0 \;\;\forall \tilde{c} \in C_k, \tilde{c} > c.$$

Hence define

$$\underline{c}_k \equiv \min\left\{ c \in C_k : \left.\frac{\partial}{\partial c'} U(\psi_m, \phi(c, c'))\right|_{c'=c} \leq 0 \right\},$$

and if this value exists let $C_k^{**} \equiv [\underline{c}_k, \chi_{k+1}/2]$; otherwise let $C_k^{**} = \emptyset$. Since for $c' = c$ we have $U_m(\phi(c)) = U_m(\phi(c, c'))$, it follows that for all $c \in C_k^{**}$, $U_m(\phi(c)) \geq U_m(\phi(c, c'))$

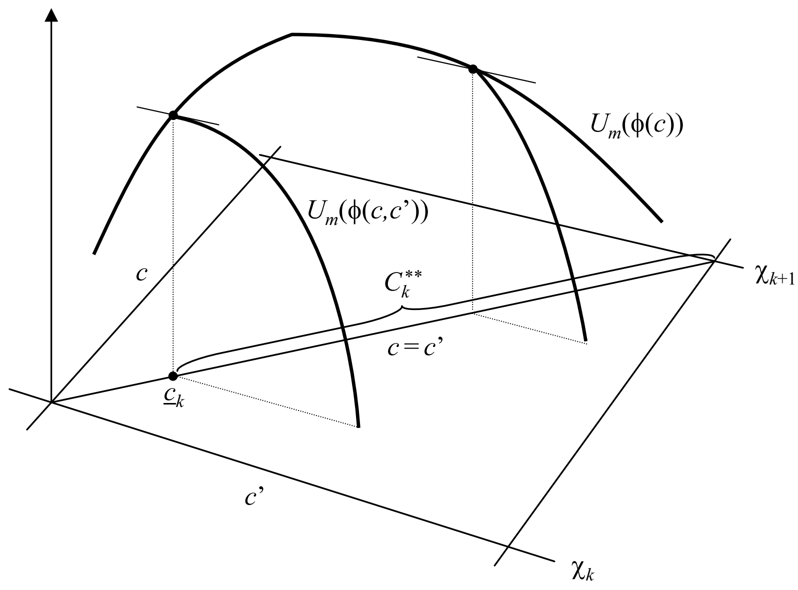

Figure 4: Construction of the point $\underline{c}_k$

$\forall c' > c$ ($c' \in C_k$). Conversely, by definition of $\underline{c}_k$, if $c \in C_k \backslash C_k^{**}$ there exists $c' > c$ such that $U_m(,\phi(c)) < U_m(\phi(c,c'))$. Hence for $c \in C_k$ to be an equilibrium it is necessary that $c \in C_k^{**}$. Figure 4 illustrates the construction of the point $\underline{c}_k$ graphically.

Step 4b. Now assume $C_k^{**} \neq \emptyset$ and take an assembly $\phi(c)$ with $c \in C_k^{**}$. Suppose that m deviates upward to some $c'' \in C_{k''}$, $k'' > k$. The expected utility for m after the deviation can be written as

$$U_m(\phi(c,c'')) = \sum_{i>k'} u((2c''-1)\psi_m) + \sum_{m<i\leq k''} u(\psi_i - \psi_m) + u((1-c'')\psi_m)$$
$$+ \sum_{i \in d(c)-m} u((1-c)\psi_m) + \sum_{i:0<\psi_i<\psi_m} u(\psi_i - \psi_m) + \sum_{i:\psi_i \leq 0} u(-\psi_m), \quad (15)$$

and the utility gain from the deviation (i.e. the difference between (15) and (9)) is therefore

$$\Delta(c,c'') = \sum_{i>k''} \left[u((2c''-1)\psi_m) - u((2c-1)\psi_m)\right]$$
$$+ \sum_{k<i\leq k''} \left[u(\psi_i - \psi_m) - u((2c-1)\psi_m)\right] + \left[u((1-c'')\psi_m) - u((1-c)\psi_m)\right]. \quad (16)$$

Holding the deviation c'' fixed and taking the derivative of $\Delta(c,c'')$ with respect to c, we have

$$\begin{aligned}\frac{\partial}{\partial c}\Delta(c,c'') &= -2\psi_m \sum_{i>k} u'((2c-1)\psi_m) + \psi_m u'((1-c)\psi_m) \\ &= -\frac{\partial}{\partial c'}U_m(\phi(c,c')\Big|_{c'=c} \geq 0,\end{aligned}$$

where the inequality follows from the fact that $c \geq \underline{c}_k$. That is, if deviating from c to c'' is profitable for m (i.e. $\Delta(c,c'') > 0$) and we take $\tilde{c} \in C_k$, $\tilde{c} > c$, then deviating from $\tilde{c}$ to c'' is profitable as well. Equivalently, if there does not exist $c'' \in C_{k''}$ such that a deviation from c to c'' is profitable, then for any $\tilde{c} \in C_k$, $\tilde{c} < c$, there does not exist such c''. Define

$$\overline{c}_k \equiv \min_{k''>k} \max \left\{ c \in C_k : \Delta(c,c'') \leq 0 \; \forall c'' \in C_{k''} \right\},$$

and if this value exists let $C_k^{***} \equiv [\chi_k/2, \overline{c}_k]$; otherwise let $C_k^{***} = \emptyset$. Hence for $c \in C_k$ to be an equilibrium it is necessary that $c \in C_k^{***}$.

Step 4c. Now let

$$C_k^* \equiv C_k^{**} \cap C_k^{***};$$

note that C_k^* is either the interval $[\underline{c}_k, \overline{c}_k]$ or $C_k^* = \emptyset$, and that C_k^* is closed because both C_k^{**} and C_k^{***} are closed. Since we have exhausted all possible deviations, for $c \in C_k$ to be an equilibrium it is necessary and sufficient that $c \in C_k^*$.

Step 5: Equilibria in C_k ($k > \kappa$)

Pick any interval $C_k = [\chi_k/2, \chi_{k+1}/2]$ with $k > \kappa$. Define C_k^* as before. The characterization of C_k^* differs from the approach in Step 4 in two aspects: First, since m is the only member of $d(c)$, m can not only enlarge the set $Q(\phi(c))$ by deviating upward but also shrink the constraint set by deviating downward, so we need to check both types of deviation. Second, it is then no longer guaranteed that if m does not want to deviate, no other agent $i < m$ would like to deviate. However, since ϕ_m is the only pivotal representative, for $\phi(c)$ to be an equilibrium all districts $i \neq m$ must elect $\phi_i^* = \psi_i$ by Lemma 2 (c), but this is the case in $\phi(c)$ when $|d(c)| = 1$. Thus, as in Step 4 above, we only need to consider deviations by the median district m. Making an argument analoguos to the one provided at the beginning of Step 4, we can further restrict the set of possible deviations and consider only those deviations from c to $\min\{\chi_{m-1}, 1/2\} \leq c' \leq 1$. Therefore, c is an equilibrium if and only

$$U_m(\phi(c)) \geq U_m(\phi(c,c')) \;\; \forall c' \text{ s.t. } \min\{\chi_{m-1}, 1/2\} \leq c' \leq 1.$$

But observe that if $\min\{\chi_{m-1}, 1/2\} \leq c' \leq 1$, then $U_m(\phi(c,c')) = U_m(\phi(c'))$. Therefore $c \in [\min\{\chi_{m-1}, 1/2\}, 1]$ is an equilibrium if and only if it maximizes $U_m(\phi(c))$ on this set. By (9), $U_m(\phi(c))$ is concave within each of the intervals $C_{\kappa+1}, \ldots, C_K$. The set of values $c \in C_k$ that maximize $U_m(\phi(c))$ on C_k, call it C_k^{**}, is thus a closed subinterval of C_k. The set C_k^* must then be either empty (in case m has a profitable deviation to $c \in C_{k'}$, $k' \neq k$, or $C_k^* = C_k^{**}$.

Step 6: Equilibria in C_κ

We have so far shown that for all $k \neq \kappa$, C_k^* is either empty or a closed interval. If $\kappa \geq m$ then there will be an interval C_κ that will contain points $c \leq \chi_{m-1}$ (for which $|d(c)| > 1$ so that only upward deviations by m need to be considered) as well as points $c > \chi_{m-1}$ (for which $|d(c)| = 1$ so that upward and downward deviations by m must be considered). For this interval C_κ, define $\underline{c}_\kappa$ and $\overline{c}_\kappa$ as in Step 4. If $\overline{c}_\kappa \leq \chi_{m-1}$ we are done, and C_κ^* is either empty or $C_\kappa^* = [\underline{c}_\kappa, \overline{c}_\kappa]$. Similarly, if $\underline{c}_\kappa > \chi_{m-1}$ we are done, since then the only possible equilibria in C_κ must be such that $|d(c)| = 1$, for which case we can go to Step 5 and conclude that C_κ^* is either empty or a closed interval. If $\underline{c}_\kappa \leq \chi_{m-1} < \overline{c}_\kappa$, however, the points in $[\underline{c}_\kappa, \overline{c}_\kappa]$ that lie to the right of χ_{m-1} may not be equilibria, as they must be checked against downward deviations as well (i.e. $c \in [\underline{c}_\kappa, \overline{c}_\kappa]$ is only a necessary condition for $\phi(c)$ to be an equilibrium).

So assume this latter case, and take $\chi_{m-1} < c \leq \overline{c}_\kappa$. Arguing as in Step 5, the only downward deviations to consider are those for district m from c to $\chi_{m-1} \leq c' < c$. However, for all $c' \geq \chi_{m-1}$, $\phi(c') = \phi(c, c')$ by definition of $\phi(\cdot)$, and thus $U_m(\phi(c')) = U_m(\phi(c, c'))$. Thus $U_m(\phi(c))$ is non-increasing in c on $[\chi_{m-1}, \overline{c}_\kappa]$ since these values satisfy the necessary conditions that upward deviations are not profitable. Hence the set of all values $c \in [\chi_{m-1}, \overline{c}_\kappa]$ which is also immune to downard deviations by m must be the interval $[\chi_{m-1}, b]$, where $b = \max\{c \in [\chi_{m-1}, \overline{c}_\kappa] : U_m(\phi(c)) = U_m(\phi(\chi_{m-1}))\}$, and thus $C_\kappa^* = [\underline{c}_\kappa, b]$.

Step 7: B is non-empty

Take $\alpha \equiv \inf\{c \in [1/2, 1] : |d(c)| = 1\}$. α exists and will be equal to χ_{m-1} if $\chi_{m-1} \geq 1/2$, and equal to 1/2 otherwise. $\phi(c)$, $c \in [\alpha, 1/2]$, will be an equilibrium if and only if it maximizes $U_m(\phi(c))$ on $[\alpha, 1/2]$, as argued in Step 5 above. Since $c \in [\alpha, 1/2]$ is compact and U_m is continuous in $\phi(c)$ and $\phi(c)$ is continuous in c, $U_m(\phi(c))$ must have a maximum on $[\alpha, 1/2]$. Let $c^* \in [\alpha, 1/2]$ be any point that maximizes $U_m(\phi(c))$; then $c^* \in C_k^*$ for some k by definition of C_k^*, and hence $B = \cup_k C_k^* \neq \emptyset$. □

Proof of Theorem 3

Step 1: Proof of claim (a)

Let $\phi = (\psi_1, \ldots, \psi_N)$. By Lemma 2 (b) there cannot be an equilibrium assembly ϕ such that $|\phi_i| > |\psi_i|$ for some i. Thus, we simply need to show that ϕ is not an equilibrium when $\chi_N > 2$. Note that $\overline{Q}(\phi) = 2\psi_m$, and since $\chi_N = \psi_N/\psi_m > 2$, $\phi_N > \overline{Q}(\phi)$. This implies that $z_N(\phi) = \overline{Q}(\phi) = 2\psi_m$. By assumption, $\psi_{m-1} < \psi_m$, so the only pivotal representative is ϕ_m, and in particular district m can decrease $\overline{Q}(\phi)$ down to $2\psi_{m-1}$ by electing representative $\phi'_m < \psi_m$. Now observe that (5) implies that the marginal change in m's expected utility from a decrease in ϕ_m is $-|u'(0)| + |2u'(\psi_m)| > 0$ since u is decreasing and convex. Hence a slight deviation to $\phi'_m < \psi_m$ increases m's expected utility, and ϕ cannot be an equilibrium.

Step 2: Proof of claim (b)

Suppose ϕ is a conservative equilibrium, so there is at least one district, say i, who elects $|\phi_i| < |\psi_i|$. By Lemma 2 (b) $0 < \phi_i < \psi_i$, and by Lemma 2 (c) ϕ_i is pivotal. We begin with a few observations:

First note that $i \geq m$ and thus $\psi_i \geq \psi_m$. To see this, suppose to the contrary that $i < m$ and thus $\phi_i < \psi_i < \psi_m$. Then $\overline{Q}(\phi) = 2\phi_m = 2\psi_m > 2\phi_i$, contradicting that ϕ_i is pivotal. Note further that $\phi_i < \psi_N/2$. To see this, suppose otherwise. Then $\overline{Q}(\phi) = 2\phi_i \geq \psi_N$, and (again using Lemma 2 (b)) no representative to the right of ϕ_i would be constrained. This implies that the marginal change in i's expected utility from increasing ϕ_i is $|u'(\psi_i - \phi_i)| > 0$, and thus a slight deviation by i to $\phi_i' > \phi$ increases i's expected utility, so we conclude that $\phi_i < \psi_N/2$. Since $\chi_N < 1 + 1/N = 2m/N$, we thus have the following inequality:

$$\phi_i < \frac{1}{2}\psi_N < \frac{1}{2}\left(1 + \frac{1}{N}\right)\psi_m = \frac{m}{N}\psi_m. \tag{17}$$

We are now ready to show that ϕ cannot be an equilibrium. Suppose that i increases ϕ_i to $\phi_i' = \psi_i$; call the resulting assembly $\phi' = (\phi_{-i}, \psi_i)$. By Lemma 2 (c) ϕ_i is pivotal in ϕ, so the deviation to $\phi_i' = \psi_i$ increases $\overline{Q}(\phi) = 2\phi_i$ to $\overline{Q}(\phi') = q > 2\phi_i$. Since by Lemma 2 (b) $\phi_i \geq \psi_i/2$, $\psi_i \in Q(\phi) \subset Q(\phi')$. The deviation to ϕ_i' has two effects on i's expected utility. First, there is a utility gain as a result of the fact that after the deviation i's representative proposes $z_i(\phi') = \psi_i$. This gain is

$$\Delta^+ \equiv u(0) - u(\psi_i - \phi_i) > u(0) - u\left(\psi_m - \frac{m}{N}\psi_m\right) = u(0) - u\left(\frac{m-1}{N}\psi_m\right),$$

where the inequality follows from (17), $\psi_i \geq \psi_m$, and u decreasing.

Second, there is a utility loss as a result of relaxing the constraint on representatives for which $\phi_j > \overline{Q}(\phi)$ which are unable to implement their most preferred policies in ϕ. The worst case for i is that all representatives $\phi_j > \overline{Q}(\phi)$ are able to implement their most preferred policies in ϕ'; the utility loss for i, Δ^-, must therefore satisfy the following inequality:

$$\begin{aligned}\Delta^- \leq \sum_{j:\phi_j > \overline{Q}(\phi)} \left[u(2\phi_i - \psi_i) - u\left(\psi_j - \psi_i\right)\right] &< (m-1)\left[u(2\phi_i - \psi_i) - u(\psi_N - \psi_i)\right] \\ &< (m-1)\left[u(0) - u\left(\frac{N+1}{N}\psi_m - \psi_m\right)\right] = (m-1)\left[u(0) - u\left(\frac{1}{N}\psi_m\right)\right].\end{aligned}$$

The second inequality follows from $\overline{Q}(\phi) = 2\phi^{(m)}$ and the fact that $\phi_j \leq \psi_j \leq \psi_N$ for all $\phi_J > \overline{Q}(\phi)$, and the third inequality follows from $\chi_N < 1 + \frac{1}{N}$ and u decreasing. Both Δ^+ and Δ^- are independent of the origin of u, so assume without loss of generality that $u(0) = 0$. Then by the concavity of u we have

$$\begin{aligned}\Delta^+ - \Delta^- &> -u\left(\frac{m-1}{N}\psi_m\right) + (m-1)u\left(\frac{1}{N}\psi_m\right) \\ &\geq -(m-1)u\left(\frac{1}{N}\psi_m\right) + (m-1)u\left(\frac{1}{N}\psi_m\right) = 0.\end{aligned}$$

Therefore, ϕ cannot be an equilibrium. $\square$

Proof of Theorem 4

Step 1: Proof of claim (a)

Fix some $\frac{1}{2} < c < 1$. That $c \geq \frac{1}{2}$ follows from Lemma 2 (b). To check if $\phi(c)$ is an equilibrium it suffices (by Lemma 1) to check that

$$U_m(\phi(c)) \geq U_m(\phi(c, c')) \ \ \forall c < c' \leq 1, \tag{18}$$

where we define $\phi(c, c') \equiv (\phi_{-m}(c), \phi_m(c'))$. With linear utility checking that (18) is satisfied involves checking only that m does not want to deviate from c to c'. To see that m does not want to deviate, let $c' \in [c, 1]$ maximize $U_m(\phi(c, c'))$ (if there are several maximizers pick the largest one). If $c' = 1$ then $\phi(c)$ cannot be an equilibrium. If $c' < 1$, then $\overline{Q}(\phi(c, c')) = 2c'\psi_m$ and for c' to be a best response for m it must be that $\psi_N > 2c'\psi_m$, i.e. at least one representative in $\phi(c, q)$ is constrained. But then, since $u'(x) = -1$,

$$\frac{\partial}{\partial c'} U_m(\phi(c, c') \leq \psi_m - 2\psi_m = 1\psi_m < 0,$$

and thus decreasing c' slightly would increase $U_m(\phi(c, c'))$. Hence $c' \in \{c, 1\}$, and so $\phi(c)$ is an equilibrium if $U_m(\phi(c)) \geq U_m(\phi(c, 1))$.

Let us define $k(a) \equiv \min\{i : \chi_i \geq a\}$ and $\overline{k} \equiv \max\{i : \chi_i \leq 2\}$. Let $d(c)$ be defined as in the proof of Theorem 2. The condition that m does not want to deviate from c to 1 is

$$\begin{aligned}
U_m(\phi(c)) &= -\sum_{i:\psi_i < c\psi_m} (\psi_m - \max\{0, \psi_i\}) - \sum_{i \in d(c)} (1-c)\psi_m \\
&\qquad - \sum_{i:\psi_i > \psi_m} (\min\{2c\psi_m, \psi_i\} - \psi_m) \\
&\geq -\sum_{i:\psi_i < c\psi_m} (\psi_m - \max\{0, \psi_i\}) - \sum_{i \in d(c)\setminus m} (1-c)\psi_m \\
&\qquad - \sum_{i:\psi_i > \psi_m} (\min\{2\psi_m, \psi_i\} - \psi_m) \\
&= U_m(\phi(c, 1)).
\end{aligned} \tag{19}$$

After some algebra (19) can be written as

$$1 - c \leq \beta(c) \equiv \sum_{i=k(2c)}^{\overline{k}} (\chi_i - 2c) + 2(N - \overline{k})(1 - c). \tag{20}$$

Thus we need to show that (20) implies $1 - r \leq \beta(r)$ for $r < c$. If $\chi_N > 2$ this must be the case since then $\overline{k} < N$ and so $\beta(r) \geq 2(1 - r) > 1 - r$. If $\chi_N \leq 2$ then $\overline{k} = N$ and (20) can only hold if $N - k(2c) \geq 0$, in which case we can write (20) as

$$1 - c \leq \sum_{i=k(2c)}^{N} (\chi_i - 2c) = \sum_{i=k(2c)}^{N} \chi_i \; - \; 2c(N - k(2c) + 1). \tag{21}$$

The right-hand side of (21) is a piecewise linear continuous function of c and has slope $-2(N - k(2c) + 1) \leq -2$ where differentiable. The left-hand side of (21) has slope -1. Therefore inequality (21) remains intact when c is decreased, so $\phi(r)$ is an equilibrium for all $\frac{1}{2} \leq r \leq c$.

Step 2: Proof of claim (b)

Suppose first $\chi_N > 2$. Then by Theorem 3 there are only conservative equilibria. But when $\chi_N > 2$ it is also true the

$$\overline{\chi} > \frac{1}{m-1}\left(2 + (m-2)\right) = \frac{m}{m-1} = \frac{N+1}{N-1} > \frac{N}{N-1},$$

the desired inequality. Suppose now $\chi_N < 2$. By part (a), it is sufficient to check whether $\phi(1/2)$ is an equilibrium, which (by the argument given in the proof of part (a)) requires only that m does not want to deviate from $c = 1/2$ to $c' = 1$. The set of feasible policies in $\phi(1/2)$ is $Q(\phi(1/2)) = [0, \psi_m]$, and the set of feasible policies in $\phi(1)$ is $Q(\phi(1)) = [, 2\psi_m]$ which contains $\phi_i = \psi_i \ \forall i > m$ due to $\chi_N < 2$. Thus $\phi(1/2)$ is an equilibrium if and only if

$$\begin{aligned} U_m(\phi(1/2) &= -\sum_{i=1}^{m-1}(\psi_m - \max\{0, \phi_i(1/2)\}) - \frac{1}{2}\psi_m \\ &\geq -\sum_{i=1}^{m-1}(\psi_m - \max\{0, \phi(1/2)\}) - \sum_{i=m+1}^{N}(\psi_i - \psi_m) \\ &= U_m(\phi(c, 1)), \end{aligned} \tag{22}$$

which can be simplified to $\sum_{i=m+1}^{N}(\psi_i - \psi_m) \geq \frac{1}{2}\psi_m$. After rearranging we get

$$\sum_{i=m+1}^{N} \psi_i \geq \left(m - \frac{1}{2}\right)\psi_m,$$

and dividing both sides by $(m-1)\psi_m$ yields the desired inequality,

$$\overline{\chi} \geq \frac{m - \frac{1}{2}}{m-1} = \frac{N}{N-1}.$$

□

Proof of Theorem 5

Proof of claim (a)

Take any $c^* \in B^*$, and any $c \in B$ such that $c > \max B^*$. Let $C = \{i_0, \ldots, m\}$ where $i_0 = \min\{i : \psi_m \geq \psi_i \geq c^*\psi_m\}$ is the set of districts electing conservative representatives in $\phi(c^*)$. We show that $\phi(c)$ is not a CPN by identifying a credible deviation by the group C.

Since $c > \max B^*$, $U_m(\phi(c^*)) > U_m(\phi(c))$. Lemma 1 then implies $U_i(\phi(c^*)) > U_i(\phi(c))$ for all $i < m$. Thus on the restricted set of players C, $\phi(c) > \phi(c^*)$ is pareto-dominated by $\phi(c^*)$, so every member of the coalition C gains can from a joint deviation to $\phi_i = c^*\psi_m$ $\forall i \in C$, resulting in assembly $\phi(c^*)$. We now establish that this deviation is also credible. To do so, take $C' \subseteq C$, and suppose all $j \in C'$ deviated from $\phi_j = c^*\psi_m$ to ϕ'_j. Call the resulting assembly ϕ'. We show that if this second-order deviation is credible, $U_j(\phi(c^*)) >$

$U_j(\phi')$ for all $j \in C'$. Hence no subcoalition C' of the initial coalition C deviates from $\phi(c^*)$, which implies that the initial deviation to $\phi_i(c^*)$ for $i \in C$ is credible. This, in turn, means that all order-preserving equilibria $\phi(c)$, $c > c^*$, are not immune to a credible coalitional deviation to $\phi(c^*)$.

For credibility it is necessary that $(\phi'_j)_{j \in C'}$ is a Nash equilibrium in the game induced on C' by holding $\phi_{-C'}(c^*)$ fixed. Then, by following the same arguments as in the proof of Lemma 2, $\phi'_j \leq \psi_j$ and ϕ'_j is pivotal in ϕ' for all $j \in C'$. This implies that $\phi'_j = \phi'_k$ $\forall j, k \in C'$. Furthermore the deviation can only be upward (otherwise pivotalness requires $C' = C$, which is impossible since $m \in C$ would lose from the deviation by definition of c^*). So $\phi'_j > c^* \psi_m$, and we define $c' = \phi'_j / \psi_m > c^*$. To show that $U_j(\phi(c^*)) > U_j(\phi')$ for all $j \in C'$, we compare three assemblies: $\phi(c^*)$, ϕ', and $\phi(c')$. Let $\overline{Q} = 2c^* \psi_m$ and $\overline{Q}' = 2c' \psi_m$. Now take $j \in C'$ and write j's expected payoff in assembly $\phi(c^*)$ as

$$U_j(\phi(c^*)) = \sum_{i:\psi_i < c^* \psi_m} u(\psi_j - \max\{0, \psi_i\}) + \sum i \in C u(\psi_j - c^* \psi_m) + \sum_{i:\psi_m < \psi_i \leq \overline{Q}} u(\psi_i - \psi_j) + \sum_{i:\psi_i > \overline{Q}} u(\overline{Q} - \psi_j). \quad (23)$$

Similarly, in $\phi(c')$, j's expected utility can be written as

$$U_j(\phi(c')) = \sum_{i:\psi_i < c^* \psi_m} u(\psi_j - \max\{0, \psi_i\}) + \sum i \in C u(\psi_j - \min\{\psi_i, c' \psi_m)\} + \sum_{i:\psi_m < \psi_i \leq \overline{Q}'} u(\psi_i - \psi_j) + \sum_{i:\psi_i > \overline{Q}'} u(\overline{Q}' - \psi_j), \quad (24)$$

and in ϕ' it is

$$U_j(\phi') = \sum_{i:\psi_i < c^* \psi_m} u(\psi_j - \max\{0, \psi_i\}) + \sum i \in C \backslash C' u(\psi_j - c^* \psi_m) + \sum i \in C' u(\psi_j - c' \psi_m) + \sum_{i:\psi_m < \psi_i \leq \overline{Q}'} u(\psi_i - \psi_j) + \sum_{i:\psi_i > \overline{Q}'} u(\overline{Q} - \psi_j). \quad (25)$$

The structure of the terms in (23)–(25) is similar to that of (9).

By Lemma 1, since $j \in C' \subseteq C$, $U_j(\phi(c^*)) > U_j(\phi(c'))$:

$$\begin{aligned} U_j(\phi(c^*)) - U_j(\phi(c')) &= \sum_{C \backslash C'} \left[u(\psi_j - c^* \psi_m) - u(\psi_j - \min\{\psi_i, c' \psi_m\}) \right] \\ &+ \sum_{i \in C'} \left[u(\psi_j - c^* \psi_m) - u(\psi_j - c' \psi_m) \right] \\ &+ \sum_{i:\overline{Q} < \psi_i \leq \overline{Q}'} \left[u(\overline{Q} - \psi_j) - u(\psi_i - \psi_j) \right] \\ &+ \sum_{i:\psi_i > \overline{Q}'} \left[u(\overline{Q} - \psi_j) - u(\overline{Q}' - \psi_k) \right] \ > \ 0, \end{aligned} \quad (26)$$

where the fact was used that $\psi_j \geq c'\psi_m$ for $j \in C$. Now note that

$$\begin{aligned} U_j(\phi(c^*)) - U_j(\phi') &= \sum_{i \in C'} \left[u(\psi_j - c^*\psi_m) - u(\psi_j - c'\psi_m)\right] \\ &+ \sum_{i:\overline{Q} < \psi_i \leq \overline{Q}'} \left[u(\overline{Q} - \psi_j) - u(\psi_i - \psi_j)\right] \\ &+ \sum_{i:\psi_i > \overline{Q}'} \left[u(\overline{Q} - \psi_j) - u(\overline{Q}' - \psi_k)\right], \end{aligned} \tag{27}$$

so that the difference between (26) and (27) is just

$$\sum_{C \backslash C'} \left[u(\psi_j - c^*\psi_m) - u(\psi_j - \min\{\psi_i, c'\psi_m\})\right] < 0,$$

where this inequality follows from the fact that $c^*\psi_m \leq \min\{\psi_i, c'\psi_m\} < \psi_j$ for $j \in C'$ and $i \in C \backslash C'$. Thus we conclude that

$$U_j(\phi(c^*)) - U_j(\phi') > U_j(\phi(c^*)) - U_j(\phi(c')) > 0,$$

as required.

Proof of claim (b)

Take any $c^* \in B^*$. $\phi(c^*)$ is coalition-proof if no coalition of districts containing $D \subseteq I$ can benefit from a credible group deviation $(\phi'_j)_{j \in D}$. For credibility it is necessary that $(\phi'_j)_{j \in D}$ is a Nash equilibrium in the game induced on D by holding $\phi_{-D}(c^*)$ fixed. Applying the same arguments as in the proof of Lemma 2 then shows that for $j \in D \backslash C$, $\phi'_j = \psi_j = \phi_j(c^*)$, so that any such deviation is equivalent to one carried out only by members of the coalition $C \cap D$. But we have showed in the proof of part (a) that no such deviation from $\phi(c^*)$ can occur, so $\phi(c^*)$ is a coalition-proof order-preserving equilibrium.

References

Alesina, A. and V. Grilli (1992): "The European Central Bank: Reshaping Monetary Politics in Europe," in: M. Canzoneri, V. Grilli, and P. Masson (eds.) *Establishing a Central Bank: Issues in Europe and Lessons from US*, Cambridge University Press, UK.

Alesina, A. and D. Rodrik (1994): "Distributive Politics and Economic Growth," *Quarterly Journal of Economics*, **109**, 465–190.

Alesina, A. and H. Rosenthal (1996): "A Theory of Divided Government," *Econometrica*, **64**, 1311–1341.

Austen-Smith, D. (1984): "Two-Party Competition with Many Constituencies," *Mathematical Social Sciences*, **7**, 177–198.

Austen-Smith, D. (1986): "Legislative Coalitions and Electoral Equilibrium," *Public Choice*, **50**, 185–210.

Banks, J. and J. Duggan (2000): "A Bargaining Model of Collective Choice," *American Political Science Review*, **94**, 73–88.

Baron, D. and J. Ferejohn (1989): "Bargaining in Legislatures," *American Political Science Review*, **83**, 1181–1206.

Bartels, L. (2002): "Economic Inequality and Political Representation," mimeo, Princeton University.

Battaglini, M. and S. Coate (2005): "Inefficiency in Legislative Policy-Making: A Dynamic Analysis," mimeo, Princeton University.

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

Bernheim, D., A. Rangel, and L. Rayo (2004): "The Power of the Last Word in Legislative Policy Making," forthcoming, *Econometrica*.

Besley, T. and S. Coate (1997): "An Economic Model of Representative Democracy," *Quarterly Journal of Economics*, **112**, 85–114.

Besley, T. and S. Coate (2003): "Centralized versus Decentralized Provision of Local Public Goods: A Political Economy Approach," *Journal of Public Economics*, **87**, 2611–2637.

Bratton, K. (2002): "The Effect of Legislative Diversity on Agenda Setting: Evidence from Six State Legislatures," *American Politics Research*, **30**, 115–142.

Brock, W. and S. Magee (1978): "The Economics of Special Interest Politics: The Case of the Tariff," *American Economic Review Proceedings*, **68**, 246–250.

Cai, H. (2000): "Bargaining on Behalf of a Constituency," *Journal of Economic Theory*, **92**, 234–273.

Chari, V., L. Jones, and R. Marimon (1997): "The Economics of Split-Ticket Voting in Representative Democracies," *American Economic Review*, **87**, 957–976.

Cox, G. (1997): *Making Votes Count*, Cambridge University Press, UK.

Eraslan, H. (2002): "Uniqueness of Stationary Equilibrium Payoffs in the Baron-Ferejohn Model," *Journal of Economic Theory*, **103**, 11–30.

Eraslan, H. and A. Merlo (2002): "Majority Rule in A Stochastic Model of Bargaining," *Journal of Economic Theory*, **103**, 31–48.

Erikson, R. (1978): "Constituency Opinion and Congressional Behavior: A Reexamination of the Miller-Stokes Representation Data," *American Journal of Political Science*, **22**, 511–535.

Fernandez, R. and D. Rodrik (1991): "Resistance to Reform: Status Quo Bias in the Presence of Individual-Specific Uncertainty," *American Economic Review*, **81**, 1146–1155.

Gans, J. and M. Smart (1996): "Majority Voting with Single-Crossing Preferences," *Journal of Public Economics*, **59**, 219–237.

Messner, M. and M. Polborn (2004): "Voting on Majority Rules," *Review of Economic Studies*, **71**, 115–132.

Miller, W. and D. Stokes (1963): "Constituency Influence in Congress," *American Political Science Review*, **57**, 45–56.

Morelli, M. (1999): "Demand Competition and Policy Compromise in Legislative Bargaining," *American Political Science Review*, **93**, 809–820.

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

Osborne, M. and A. Slivinski (1996): "A Model of Political Competition with Citizen-Candidates," *Quarterly Journal of Economics*, **111**, 65–96.

Persson, T. and G. Tabellini (1994): "Representative Democracy and Capital Taxation," *Journal of Public Economics*, **55**, 53–70.

Persson, T., G. Roland, and G. Tabellini (1998): "Toward Micropolitical Foundations of Public Finance," *European Economic Review*, **42**, 685–694.

Persson, T., G. Roland, and G. Tabellini (2000): "Comparative Politics and Public Finance," *Journal of Political Economy*, **108**, 1121–1161.

Poole, T., and H. Rosenthal (1991): "Patterns of Congressional Voting," *American Journal of Political Science*, **35**, 228–278.

Redoano, M. and K. Scharf (2004): "The Political Economy of Political Centralization: Direct versus Representative Democracy," *Journal of Public Economics*, **88**, 799–817.

Rodrik, D. (1996): "Understanding Economic Policy Reform," *Journal of Economic Literature*, **34**, 9–41.

Rogoff, K. (1985): "The Optimal Degree of Commitment to an Intermediate Monetary Target," *Quarterly Journal of Economics*, **100**, 1169–1190.

Romer, T. and H. Rosenthal (1978): "Political Resource Allocation, Controlled Agendas and the Status Quo," *Public Choice*, **33**, 27–43.