University of Leicester

Department of Economics

Party Formation and Competition

Daniel Ladley, University of Leicester

James Rockey, University of Leicester

Working Paper No. 10/17

May 2010

Revised September 2010, December 2013, March 2014.

Party Formation and Competition

Daniel Ladley

Department of Economics
University of Leicester*

James Rockey

Department of Economics
University of Leicester

September 2013

Abstract

In the majority of democratic political systems, districts elect representatives, who form coalitions, which determine policies. In this paper we present a model which captures this process: A citizen-candidate model with multiple policy dimensions in which elected representatives endogenously choose to form parties. Numerical analysis shows that in equilibrium this model produces qualitatively realistic outcomes which replicate key features of cross-country empirical data, including variation consistent with Duverger's law. The numbers of policy dimensions and representatives elected per district are shown to determine the number, size, and relative locations of parties. Whilst multi-member district systems are found to reduce welfare.

Keywords: Citizen-Candidate Model; Political Competition; Party Formation; Duverger's Law; Computer Simulation
Words 8362

*We are greatful to Toke Aidt, Subir Bose, James Cranch, Gianni De Fraja, Miltos Makris, Suresh Mutuswami, Ludovic Renou, and Javier Rivas for helpful comments and suggestions. We also thank seminar audiences at, CEF(London) Econometric Society European meeting (Oslo), European Political Studies Association (Reykjavik), Institut d'Economia Barcelona, Silvaplana Workshop on Political Economy, and the 4th CESifo workshop on Political Economy. An eariler version of this paper was circulated with the title 'A Model of Party Formation and Competition' Email: Ladley-dl110@le.ac.uk, Rockey-jcr12@le.ac.uk

Introduction

A key feature of modern representative democracies is that political competition is dominated by political parties. Some parties are small and ideologically cohesive, others large collections of politicians with quite different views. Moreover, the size distribution of parties varies meaningfully across democracies. This variation is important, in part because the set of possible governing coalitions and hence the set of policy outcomes is contingent on the size-distribution of parties(Hart and Kurz, 1983).

This paper studies the formation of political parties and party systems. A model is developed and simulated in which elections are contested by endogenously formed political parties. These parties are comprised of politicians from multiple districts, each with a large number of voters, each of whom have preferences over a multidimensional policy-space. By varying the dimensionality of the policy-space we find that it is a key determinant of the political landscape, and in particular that the results of the model are much closer to empirical party size distributions when we consider multidimensional policy-spaces. We find new evidence for Duverger's law, and importantly highlight the importance of the interaction between political institutions and the form of voters' preferences. We provide evidence that the predictions of the model coincide with the empirical evidence. Welfare analysis suggests that as well as affecting agents' utility through the number of parties formed, that both institutions and the overall distribution of preferences have a direct impact upon average welfare.

We begin by considering what Ostrom (1986) termed *Plott's Fundamental Equation [of Public Choice]*" Plott (1979):

$$\text{Preferences} \oplus \text{Institutions} \oplus \text{Physical Possibility} \mapsto \text{Outcomes}$$

Where $\oplus$ is an abstract operator. In the context of understanding the determination of political outcomes, the relevant preferences are those of voters, and the relevant institutions comprise (the form of) a democratic system. How voters' preferences and political institutions, separately and jointly, determine political outcomes has been the subject of a venerable mountain of academic endeavor[1] Much of this work has focused on the accurate measurement or description of Preferences and Institutions. Attempting to recover the properties of $\oplus$ has been a core objective of much work in Political Science, Social Choice, and Political Economy theory. Why the structure of political competition varies, and what the consequences of this will be has been subject to much scholarly attention. This paper employs a novel approach to provide new insight into these questions. It develops a methodology designed to bridge the gap between the insights provided by formal, yet necessarily simplified models, and the richness provided by detailed, but perhaps less general, analyses of particular party systems or contexts. We develop and simulate a rich theoretical model with three key features. Firstly, political parties are simply voluntary coalitions of elected politicians formed endogenously for mutual (electoral) benefit.[2] Secondly, a key feature of politics, in practice, is that not only are politicians themselves heterogeneous but so are the electoral districts they represent, something which we also model explicitly. Here, we study the election of candidates from an empirically realist number of heterogeneous districts. Finally, both the elected and their electors may have multidimensional preferences, that is any two of them may agree on some issues but disagree profoundly, on another, unrelated, issue.

We analyze numerically a citizen-candidate model in which candidates endogenously join and form political parties. We use a new algorithm to identify the set of equilibria for each combination of parameters and study how the distribution of outcomes defined by these equilibria depends on the parameters chosen. In particular, the parameters describe electoral rules, and the dimensionality of voters preference distributions. Computational approaches to Political Science have historically been comparatively rare. We argue that recent developments in the relevant theory alongside advances in computational power makes the numerical analysis of empirically realistic, yet theoretically grounded models productive.

Our starting point is a citizen-candidate model, as introduced by Osborne and

[1]That there are also large, multifaceted, literatures dedicated to understanding their determinants may suggest, as Plott (1979) notes, that this 'Law' is not as fundamental as it might be.

[2]In reality, political parties perform other functions. These are not studied here, partly because of pronounced national differences in the other functions of parties.

Slivinski (1996) and Besley and Coate (1997), incorporating many policy dimensions, many districts, and in which politicians may or may not choose to form or join political parties. In our setting there will be two strategic decisions, whether to stand for election, and which party to join. Voting is sincere. The first strategic decision - candidacy - follows Dutta, Jackson and Le Breton (2001) who show that strategic candidacy is necessary for voting outcomes to be regarded as strategic. Equilibrium party membership are given by a (computational analogue) of the bi-core stability notion employed by Levy (2004) and owing to Ray and Vohra (1997). Given we model these two key decisions the outcomes of the simulations may be regarded as representing strategic equilibria. But, they will in each case represent only one outcome of potentially many, and will be dependent on, as per Plott's equation, preference distributions and the institutional parameters chosen. So that the results of this approach are comparable and complimentary to, previous analytic work then we characterize the full set of (strategic) equilibria. This is done using a novel algorithm that is shown to do so with a good degree of certainty.

The model of Morelli (2004) is of particular relevance for this paper. Morelli's specific objective is to provide a framework where the *'Duvergian predictions can be studied even when the electorate is divided into multiple districts and candidates and parties are separate entities.'* He finds support for the Duvergian hypothesis – that plurality electoral systems lead to two party systems – and his setup incorporates what he claims are the necessary features of *'strategic voters, strategic parties, and strategic candidates, within and across districts'*. As will be discussed below, in his model political parties provide a means of coordinating voters within and between districts as well as a method by which coalitions of heterogeneous candidates can commit to a shared policy-platform. One contribution of this paper is to build on his insights to show that as well as the Duverger's law, a citizen-candidate framework can give rise to many other major features of contemporary democracies.

In taking a computational approach to political science, the most similar work to ours is that of Bendor, Diermeier, Siegel and Ting (2011), Laver (2005), and Kollman, Miller and Page (1997).[3] Laver (2005) also employs a computational approach to the analysis of political parties, but in a very different manner to this paper. He focuses on the dynamic properties of competition between parties, with pre-specified behavioral rules, over time. Our focus is the model's steady state and in particular the equilibrium distribution of parties as the type of electoral system and number of policy dimensions varies. Similarly, whilst both models are compared to empirical data, Laver (2005) uses time-series data to study dynamics within a system, whereas we focus on cross-national comparisons, to study variation across systems. Many of the results of this paper are obtained by simulating our model many times and analyzing the distribution, thus abstracting from any given preference distribution. In this respect our approach is more similar to Kollman, Miller and Page (1997) who use a related technique to study a Tiebout type model. Bendor, Diermeier, Siegel and Ting (2011) analyse Aspiration-Based Adapative-Rules that are designed to capture the bounded-rationality and behavioural biases of both politicians and voters. Again, our focus is different in that we are concerned with studying fully strategic outcomes and again abstract away from particular outcomes.

The paper is organized as follows. Section 2 presents the model. Section 3 details how the model is simulated. Section 4 presents results and analyses the qualitative features of the models' equilibria as well as discussing their form and number. Section 5 focuses on the relationship between the number of preference dimensions and the size distribution of parties. Section 6 shows the results of the model emulate well the empirical relationships between the number of parties, the electoral system, and the distribution of preferences. Section 7 studies the consequences for welfare of variation in institutions and preferences. Section 8 concludes.

[3] There is a small literature applying computational techniques to problems in political economy. The first paper of which we are aware to study voting is that of Tullock and Campbell (1970) who analyzed computationally the problem of cyclical majorities in small committees with multi-dimensional preferences. They found that the impact of additional preference dimensions beyond two was small. Although our setting is different, the results of our model suggest similarly that the key difference is between having one or more than one dimension. A key contribution was that of Kollman, Miller and Page (1992) who in contrast to much of the previous rational choice literature, studied the behavior of boundedly rational parties. He argued that the, sometimes incomplete, platform convergence predicted by analytic models was robust to non-fully rational parties. This type of question, involving the behavior of a large number of boundedly rational agents lends itself to simulation-based approaches.

2 Model

The model has two stages, in the first stage citizens decide whether to stand for election and vote for their most preferred candidate. Those that run may do so on the platform of a party, which may differ from their most preferred policies. In the second stage, elected citizens (politicians) consider whether to change party.

2.1 Citizen-Candidates

The model is a spatial-voting model in the tradition of Downs (1957). There is a population of J citizens split between D districts each with preferences over policy outcomes. Their utility is dependent on the distance between the point in a policy-space representing their ideal outcome and the point representing the implemented policy. We therefore begin by defining the policy-space and an associated metric. Voters have preferences defined on the N-dimensional unit hypercube, $[0,1]^N \in \mathbb{R}^N$. Individual $j \in J$ has an ideal point within this space denoted $A_j = [a_{j1},..,a_{jk},..,a_{jN}]$ where a_{jk} is their preferred point in dimension k. In keeping with much of the literature we use the Euclidean norm. Our results are robust to the alternative of the l_1 norm, and we discuss this choice given the recent work of Humphreys and Laver (2010) and Eguia (2013) in Appendix A.

The defining feature of citizen-candidate models is that any citizen may choose to stand for election. If they do then they incur a cost $\kappa \geq 0$, reflecting fiscal and psychic costs of running for election. If they manage to be elected they receive a rent of $\gamma \geq 0$. We denote citizen j's utility from an implemented policy $W[w_1,...,w_N]$ as $U^j(W)$, and the distance as follows:

$$|(W - A_j)| = \sqrt{\sum_{k=1}^{N}(w_k - a_{jk})^2} \tag{1}$$

Normalising such that $U^j(W) : [0,1]^N \mapsto [0,1]$ we have the following pay-off structure:

$$(2) \qquad U^j(W) = \begin{cases} 1 - \frac{|W-A_j|}{\sqrt{N}} - \kappa + \gamma & \text{if she is elected} \\ 1 - \frac{|W-A_j|}{\sqrt{N}} - \kappa & \text{if she is not elected} \\ 1 - \frac{|W-A_j|}{\sqrt{(N)}} & \text{if she does not stand} \end{cases}$$

Given a set of candidates, individuals vote sincerely for the candidate that has the most similar platform to their preferred policy. Thus, the strategic decision is whether or not to stand. Dutta, Jackson and Le Breton (2001) show that outcomes of all democratic voting procedures depend on the candidacy decision of those who don't (cannot) win the election in question. One of the important contributions of Morelli (2004) is that often with endogenous candidacy equilibrium rational (strategic) voting behavior is sincere.[4] But, crucially, Morelli shows that "*the equilibrium policy outcome is not affected by whether voters are expected to be sincere or strategic. Thus, the sincere versus strategic voting issue is irrelevant for welfare analysis.*"

2.2 Parties

A key feature of representative democracy is that politicians normally belong to a particular party. Sometimes such parties have extensive histories and varying positions over time, elsewhere parties come and go, split and merge, and so forth. Politicians may also change parties during their career – for example, Churchill – and the groups that a particular party represents may even be reversed over time – consider, the Democratic Party and the Southern United States. In both reality, and in this paper, parties are made up of politicians with heterogeneous preferences who are elected from a number of districts, the population of each of which has a different preference distribution.

Given that parties are so prevalent it is instructive to consider how and why parties form. We do not presuppose the existence of any parties and abstract from many of

[4]Specifically, he shows that under the plurality rule that *'equilibrium [strategic] voting behavior is always sincere'*. But that in a proportional representation system if there is no party with more than half of the votes there will always be some voters who in equilibrium vote strategically.

the diverse functions they perform. Political parties have many roles in a democracy, and a variety of these have been modeled (these are surveyed by Merlo, 2006). Dhillon (2005) include parties as representing specific constituencies or groups, (c.f. Snyder and Ting, 2002 or Roemer, 1999), or voter coordination devices (Morelli, 2004). In Osborne and Tourky (2008) parties are modelled as a cost-sharing technology. In the model of Levy (2004) parties are devices that allow politicians to credibly campaign on a platform known not to correspond to their most-favored as party membership provides a complete contracting mechanism.[5] In this paper, parties are modeled as representing both a commitment device and also a cost-sharing technology. We argue that the combination of these two technologies represents a parsimonious way to capture much of what parties do.

In the model below, this role of political parties emerges endogenously - candidates seeking re-election often stand with platforms, different from their preferred policy if this changes the implemented policy sufficiently in their favor.

Once elections have taken place in each of the D districts the set of elected representatives together determine the policy to be implemented. We make no assumption about the pre-existence, or otherwise, of coalitions with more than one member. But, newly elected candidates either start their own coalition, with size 1, or choose to join an existing coalition.[6] In the spirit of Levy (2004) and Morelli (2004), if they seek re-election, all members of a coalition are constrained to stand on their party's platform. After individuals have joined coalitions and the coalition dynamics described below have occurred, the preferred policy of the largest coalition is implemented. Here we focus on the formation of parties rather than governments and as such we don't allow the formation of multi-party coalitions post-election.[7]

For simplicity we define the preferred policy of a coalition to be the mean of the ideal

[5] When this role of political parties is important in equilibrium varies depending on the context. In Levy (2004) the commitment device provided by political parties is unimportant with one policy dimension but allows for stable equilibria to exist in the case of multiple dimensions. In Morelli (2004) the commitment technology allows parties comprised of candidates with different preferred policies to stand, but in his setting it is rarely important.

[6] Note, whilst every representative is assigned to a coalition, given that a coalition can have a membership of 1, this is equivalent to allowing individuals not to join a coalition.

[7] This process has been the subject of much study, and Dhillon (2005) provides an excellent review.

points of its members.[8] We assume that representatives employ a heuristic of the following functional form:

$$V_r^j = \frac{\#r}{\sqrt{\sum_{k=1}^{N}(w_{ik} - \mu_k)^2 + \eta^2}} \tag{3}$$

where r denotes a coalition, $\#r$ the number of members in that coalition, and μ_r is that coalition's current policy. η is a parameter taking a small positive value so that the utility of single-member coalitions is defined. This heuristic is used to determine individuals satisfaction with membership of a particular coalition. Representatives face a trade-off: membership of a larger coalition increases the likelihood that an individual's preferences will have some influence on the implemented policy. However, research suggests that individuals dislike belonging to the same party as those very ideologically distant from themselves (c.f. Baylie and Nason, 2008). Individuals trade off the increased chance of being elected with potentially sacrificing the proximity to their preferred platform. The aim of this mechanism is to implement a minimal coordination technology which provides as parsimonious as possible a representation of the benefits and costs of party membership. It is argued that this abstraction captures the key thrust of the Osborne and Tourky (2008) model of parties as a cost-sharing technology.[9] Note, that the pay-off V_r^j does not enter equation 2 directly, it only determines party membership. The affect on an individual's utility is via the policy outcome. In this way, we require that the rents from office are always γ, and that potential politicians, like all citizens, are purely policy motivated. Parties, are simply a mechanism by which individuals come together to affect policy. By, not introducing a direct effect of party membership into the utility function we ameliorate concerns that the existence of parties is preordained and only in some lesser respect an endogenous feature of the model, rather they only exist to the extent that they assist individuals in achieving their political objectives.

[8] The results are essentially unaltered if we assume instead that policy is determined by an elected party leader.

[9] In fact, modeling explicitly a cost-sharing technology *à la* Osborne and Tourky (2008), in addition to the existing preference for larger parties, does not meaningfully alter the results presented below.

The equilibrium party membership allocation is in part determined by the choice of stability concept. This choice is informed by two requirements: that the attractiveness of a a party depends on its other members, and that politicians are free to create, split, and merge parties. Levy (2004) emphasizes the first requirement and develops a stability notion that "no group of politicians wish to quit its party and form a smaller one". Levy (2002), considers both cooperative and non-cooperative alternative formalizations of this requirement. Of these, the bi-Core is particularly attractive as it is a partition of politicians into parties corresponding to the notion above but it further allows for politicians to also merge extant parties, encapsulating a second notion that "no party wishes to merge with another". Allowing for this process of agglomeration is important because we do not assume parties exist*a priori*, and we will be interested in their equilibrium size. We find this partition iteratively, but conceptually the approach is unaltered: as implemented the bi-Core stability notion requires that no pair of parties wish to merge and no subset of any party wants to split that party into two smaller parties (potentially of size 1).

Figure 1 reveals the overalls structure of the model. In each district an election takes place according to equation 1 and as described in more detail in Appendix A.1. The citizens elected in each of the D districts comprise the population of representatives who may choose to form national coalitions. Given sincere voting, the support for each of these coalitions can be represented by the area of the policy-space for which their platform is closest to voters' bliss points. These are represented by the three shaded regions in the top pane.[10] The trade-off faced by politicians at the national level between influences and ideology then influence the relative benefits of office, and thus feeds back into who stands and (who wins) at each district level election, and so on. It is this feedback mechanism that relates party membership choices to individuals' utilities.

[10]The shading is in fact a Voronoi tessellation, as in Degan and Merlo (2009). For further details see Appendix A.1.

Figure 1: From Districts to Parties

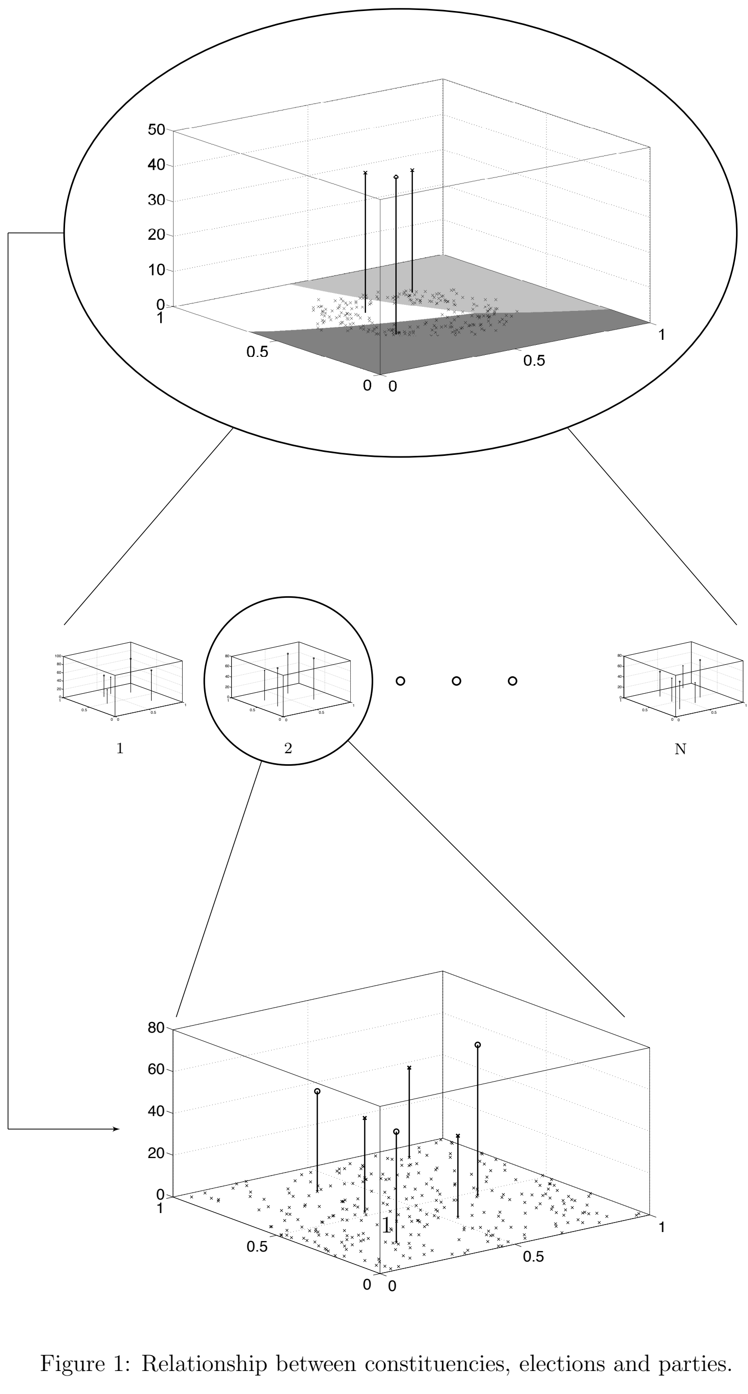


Figure 1: Relationship between constituencies, elections and parties.

3 Solving the model

The model is solved computationally as doing so analytically is infeasible. We relegate the details of how this is done to Appendix A. Here we briefly consider how the results may be interpreted. As noted above, in this model an equilibrium is defined by two choices: who stands for election and who does not; and the elected candidates choice of party. Both the choice of candidacy and the allocation of representatives to parties are in general too computationally burdensome to calculate directly. We outline the algorithm used for each stage in turn.

Most politicians in equilibrium will belong to a party and thus are committed to a platform different to the one they prefer. Hence, it is important that party memberships and platforms are the result of equilibrium behavior by the elected politicians. As discussed in the previous section, we use the bi-Core (Ray and Vohra, 1997) as our stability notion. This requires checking that no subset of individuals wishes to leave a party, and that no two parties wish to merge. Given the initial allocation of representatives to parties, we check first for subsets wishing to secede using the k-means algorithm (Lloyd, 1982) and then for each pair of parties whether both parties would have (weakly) higher utilities from merging. This process is repeated until a stable allocation is found. The resulting partition is then bi-Core stable. Further details are described in Appendix B.

The decision to stand, given an allocation of parties, is also solved using an iterative approach. Here, citizens learn their respective pay-offs from standing for elections, via experimentation. They initially stand randomly, but learn their utility from standing or not as the model is iterated. This setting is equivalent to the stimulus-response environment considered by Rustichini (1999). He shows that *"Linear procedures always converge to optimal action in the case of partial information"*. Since agents do not observe counterfactuals, inline with his results, we adopt a linear learning rule. The model is iterated until all citizens stand or not with certainty, i.e. $P_j^{stand} = \{0, 1\}$, and as described above there is a stable party membership partition.[11] This process is described

[11] As the standing probabilities become very close to zero or one, they converge increasingly slowly. Therefore, we truncate this process when $P^{stand} \pm \frac{1}{10^9} = \{0, 1\}$. Again the results are not sensitive to this assumption. Similarly, for a party membership partition to be regarded as stable we require there to

in Appendix A and is equivalent to solving for pure-strategy equilibria.[12]

3.1 Parameter Choice

The parameters determining the dimension of the policy-space and the number of representatives per district are of direct interest and are discussed below. Given a choice of these two parameters, the specific outcome of the model will depend on the random number seed chosen. This determines the distribution of voter preferences as well as the decision to stand or not before the model converges to equilibrium.[13] We are not interested in results for a given, arbitrary, seed and focus on the statistical properties obtained for a sufficiently large number of different seeds that we can be confident that the distribution of outcomes is not dependent on the seeds chosen.[14] The random seed determines preferences for citizens according to a given distribution. Our main results assume normally distributed preferences with random district mean and standard deviation. The combination of many runs and treating as many parameters as possible as random means we can be confident that the results are not contingent on a particular set of individual preferences in each constituency. Random district means and variances are used to generate realistic inter- and intra-district variation whilst also being relatively parsimonious. Whilst, for example, a uniform distribution represents a comparatively extreme assumption about the societal distribution of preferences. The results are not sensitive to this assumption, results (available upon request) were also obtained using multivariate uniform, bi-modal, as well as a normal distribution with fixed parameters.

have been no change in the membership of parties for 30 time-steps after all candidacies are known with (approximate) certainty. Again, this assumption is unimportant.

[12]The decision not to analyze mixed strategies reflects a desire to prioritize ease of interpretation and computational efficiency. The latter is important, as although the iterative approach implemented is efficient, mixed strategies are much more computationally expensive. Anecdotally, this assumption is less restrictive than it might seem. Equilibrium in mixed strategies seems to be comparatively rare: the model rarely fails to converge to candidates standing or not, with certainty. Why this is the case is a topic for future research.

[13]This latter consequence may, via path dependence, determine the choice of equilibria if there are multiple equilibria. However, as discussed in the Appendix A the consequences of such initial conditions will become negligible given a sufficiently slow learning rate.

[14]The results presented are for 1000 different seeds for each parameter combination. However, as discussed in Section B.3 of the Appendix studying an additional $50,000$ seeds identified few additional equilibria. Unfortunately, the computational demands of the model mean it is impractical to use this many different seeds for every parameter combination. As an illustration, the results for the parameters chosen require over $300,000$ CPU/hours (equivalent to around 35 years on a single modern CPU).

That results were largely unchanged. The reasons the underlying preference distribution matters little can be understood on the basis of a central limit theorem type argument that the distribution of representatives' preferences is normally distributed whatever the underlying population preference.

The results reported are for a simulated democracy in which $c = 120$ candidates are elected together representing $n = 12,000$ voters split between the C constituencies of (randomly) varying numbers of voters, each returning an equal number of representatives.[15] Larger populations may be simulated, however, this does not effect the results obtained but does dramatically increase the computational burden of the model which is of the order $\mathcal{O}(n^2)$. There are also further parameters to be chosen. The citizen-candidate model requires two parameters: γ, the rents from being elected and, κ, the cost of standing. To ensure the denominator in equation 3 is strictly positive η is added, thus avoiding the utility of singleton coalitions being undefined. Two further parameters govern citizens learning behaviour in calculating the equilibrium candidacy choices: β the rate at which citizens learn from standing versus non-standing; and $P_{j,0}^{stand}$ the initial chance of standing for election - see Appendix A for details. The results below are calculated with parameters as follows $\{\gamma = 0.2, \kappa = 0.1, \eta = 0.05, \beta = 0.99, P_{j,0}^{stand} = 0.5\}$. The behavior of the model is not contingent on any of these particular parameter values.[16]

4 Results

4.1 An example

As discussed in Section 2.2 the existence of coalitions isn't assumed *ex ante*, but they (potentially) emerge endogenously. The minimal assumptions about the benefits and costs of coalition membership give rise to stable electoral coalitions - political parties. In this section we first present some typical examples of the the joint distribution of size and

[15] The choice of 120 representatives is solely because it has many factors, but again this assumption is unimportant for the results.

[16] Results obtained for a range of alternatives are available on request. It is worth noting that the computational burden associated with obtaining results for the necessary number of runs to make inferences about the distribution of outcomes for different values of parameters prevents any attempt at 'parameter-mining' to elicit the 'best' performance of the model relative to given criteria.

relative location of these parties in equilibrium. Henceforth, we refer to such distributions as party landscapes. Whilst equilibria are still defined by individuals' standing decisions and party memberships, as discussed below, they are adequately described by the party size-location distribution up to an orthonormalization and in the remainder of the paper we treat a landscape as identifying one or more equilibrium.

We apply a Gramm-Schmidt orthonormalization (as described by Golub and Van Loan, 1996) to the party landscapes, this recasts the parties positions such that they are relative to the largest party. This has the advantage of reducing the number of dimensions needed to represent a party system, and crucially allows for comparisons of party landscapes. The details of this are described in Appendix B.

The examples in Figures 2 - 4 display the normalized landscapes on axes chosen such that the largest party is at the origin and each additional party requires an additional dimension to represent its relative location. That is, the second largest party falls on the x-axis, the third on the xy-plane etc. A further consequence is that the second party will always have a positive x-coordinate, the third party a positive y-coordinate, but potentially negative x-coordinate, and so on.

We present these results by plotting the location of parties in the first three dimensions only. Each party is represented by a sphere with diameter proportional to the number of its members. The left-hand plot depicts all 3 dimensions, the right-hand side plot shows the xy-plane. A simple example is presented in Figure 2 , with just two parties competing in a 3-dimensional policy-space with 3 candidates per district. It is worth noting that the ideological discrepancy between the parties is small, but non-zero. In general we find very few cases where there are large amounts of dispersion between the larger parties. It is argued that this is similar to the case of most mature democracies, where the main parties aren't normally extremist. Figure 3 considers an example with three parties. In this example, the additional party is to the 'left' of the two larger parties on the first dimension and also differentiates itself on the y dimension. Again, the ideological differences are relatively small. The final example displayed in Figure 4 involves five parties. Again we don't observe extremist behavior, rather the parties differentiate themselves, a little in several dimensions. The exception is the fifth party which appears relatively extreme, but in the first 3-dimensions at least, is about 20 percent of the total length of each dimension away from the largest party. Whilst a smaller, more ideologically distinct, party seems to coincide with many democracies experience. That these differences would seem in some sense to be limited is considered to be both realistic and also consistent with the central intuitions of the citizen candidate model. We present a more detailed analysis of the general relationships between district size, the number of policy dimensions, and equilibrium outcomes in the next section.

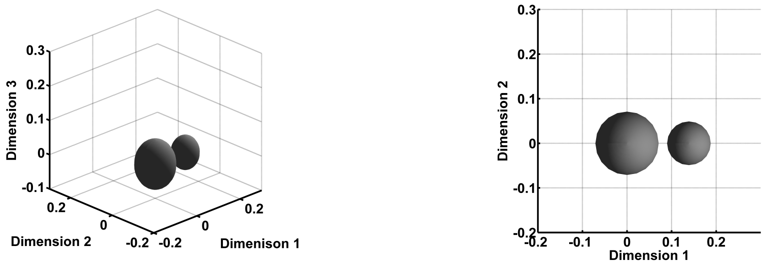


Figure 2: An example with two parties, 3 representatives per district, and 3 dimensions

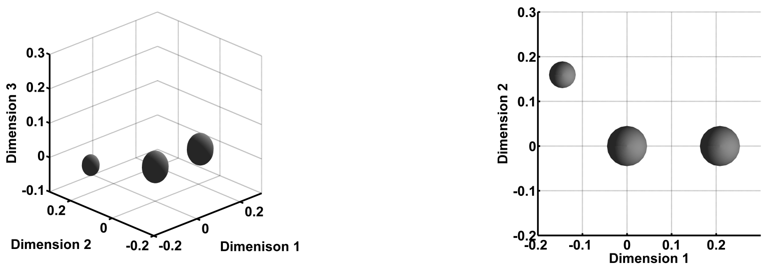


Figure 3: An example with three parties, 3 representatives per district, and 3 dimensions

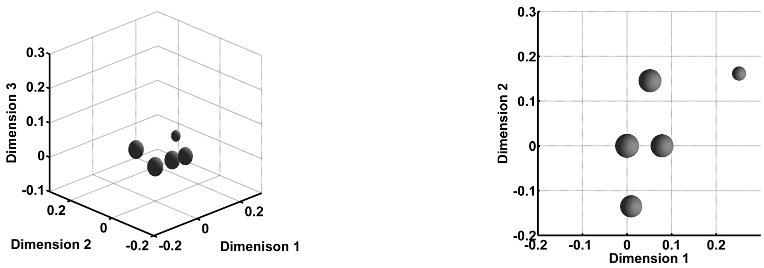


Figure 4: An example with five parties, 3 representatives per district, and 3 dimensions

It is worth considering these graphs together. They depict multi-party stable equilibria within a multidimensional policy-space, costless voting, and with intermediate degrees of polarization. In this sense they provide strong support for the use of citizen-candidate type approaches to analysing political competition and policy formation.

That the model generates qualitatively realistic results would be of limited interest if this were a one-off. More useful is to consider all of the equilibria for each combination of parameters. We are interested in the the set of distinct landscapes — What different party configurations are equilibria of the model? We apply the k-means algorithm to the orthonormalized results of the model for 1000 random number seeds. Here, the algorithm (statistically) identifies the set of distinct landscapes, but conflates those that are minutely different due to path dependence. Appendix B describes this algorithm in more detail. However, Figure 5 shows the results of this procedure applied to simulations of the case of 3 policy dimensions and 4 representatives per district. In this case we plot simultaneously the different landscapes that are equilibria of the model, where each equilibria is represented by the mid-point of the associated cluster found by the k-means algorithm.

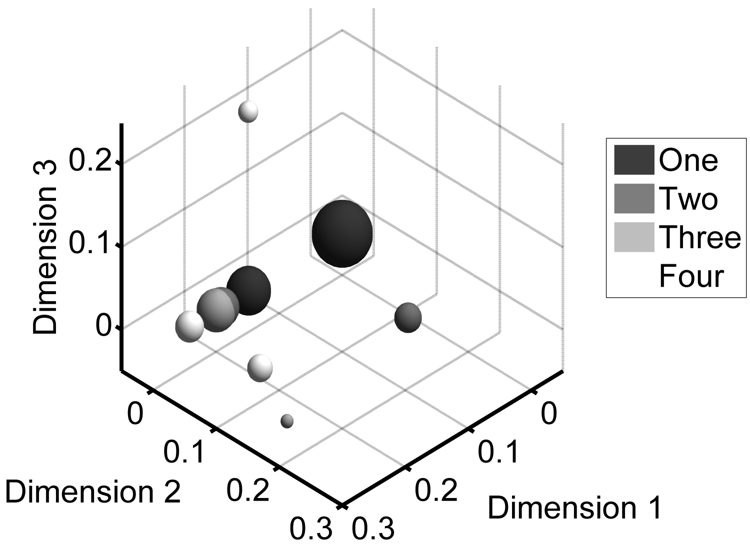


Figure 5: Four of the 9 equilibria in the case of 3 policy dimensions and 4 representatives per district. In each case the largest party is at the origin.

There is one 2-party equilibrium, two 3-party equilibria, and a 4-party equilibrium. The 2-party equilibrium, black, is straightforward to interpret and as a consequence of the Gram-Schmidt scheme the parties are only distinguished on the x-axis. The dark-gray and light-gray 3-party equilibria both contain a second large party, in virtually identical locations. The difference between the two is that the largest party is larger in the case of the light-gray equilibrium, and the third party is located further away from the origin. In the case of the dark-gray equilibrium, the third party occupies a position between the two larger parties on the x axis, and is similarly distinct on the y axis as its equivalent in the light-gray equilibrium. Again, it is argued that these outcomes all are easy to interpret intuitively. The white four-party equilibrium is perhaps less intuitive. Four way competition in 3 dimensions is inevitably complex, but the outcome with four similarly sized parties, two of which are located at the same point on the x-axis but distinguished on the y-axis, with the fourth party located equidistant from the largest two parties on the x-axis but, of course, distinguished on the z-axis. That two similarly sized parties can be similarly located and not-merge suggests that the model does not lead to parties merging too often, and similarly that there is in none of the four equilibria a tail of independent representatives suggests that similarly the parties aren't unrealistically fragmented.[17]

Note that the multiple equilibria are associated with different underlying preference distributions. One can envisage transitions from one equilibrium to another as preferences change giving rise to plausible dynamics. As mentioned in the introduction computational approaches to dynamic aspects of political competition have previously been studied by Laver and Sergenti (2011), but this would be an interesting alternative approach.

Section B.3 of the Appendix outlines results demonstrating that the algorithm successfully identifies all of the equilibria of the model. Results also suggest that the number of equilibria is decreasing in both the number of dimensions and the number of members per district.

5 The Number of Policy Dimensions

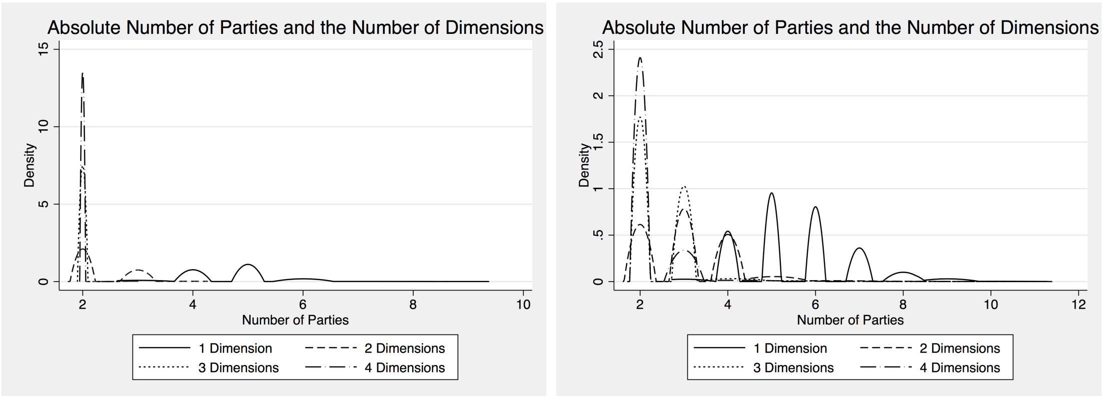
We now consider the effect of the number of policy dimensions on the distribution of equilibrium outcomes. For simplicity we initially consider the case of single-member districts. As will be shown below, the results are in fact stronger, for the case for multiple-

[17]We don't model variations in regional politics that might give rise to such parties for other reasons.

member districts.

Figure 6 displays a kernel density plot of the number of separate parties for the case of one to four policy dimensions. The pattern is clear: the number of parties is smaller when there are more dimensions. However, this effect is most marked for the transition from one to two dimensions. The addition of further dimensions beyond two is less important.

Figure 6: The number of parties decreases with the number of policy dimensions



(a) Number of Parties for Different Numbers of Dimensions, single-member districts

(b) Number of Parties for Different Numbers of Dimensions

Counting the number of distinct parties can sometimes be misleading. Not all parties are equally (electorally) important. As Laakso and Taagepera (1979) note, many democracies are characterized by a tail of small parties, which should not necessarily be given the same weight when making comparisons as larger parties as they in general have less impact on the democratic process. This tail is also a feature of the model presented here. Similarly, a party system in which there are two parties with approximately equal vote shares has in some ways more parties than one in which one party has all but two seats, which are split between two others. Hence our approach is to focus on the number of 'effective parties', N_s. This is captured by the the Laakso-Taagepera Index, the reciprocal of the sum of the party vote shares:

$$N_s = \frac{1}{\sum_{i=1}^{n} p_i^2} \tag{4}$$

Where p_i is the vote share of party i and there are n parties.[18]

Figure 7(c) is identical to Figure 6 except using the LT measure. The results confirm those for the absolute number of parties, except that it is now clearer that additional higher dimensions do impact the size distribution of parties as a whole; not, for example, the number of tail parties.

It is worth comparing our findings to those of Levy (2004). She finds that parties are 'ineffective' in single-dimensional spaces. That is they don't allow politicians to stand on platforms they don't prefer. Here, we obtain an analogous result. The distribution for $dim = 1$ in Figure 6 is qualitatively different to those for $dim > 1$. Similarly, the probability of there being one or more politicians not in a political party (in a party of size 1) is ten times as large for $dim = 1$ as it is for $dim > 1$. However, in our richer setting parties still form in the single-dimension case, one explanation for this is we consider a number of heterogeneous districts, platform commitment may be useful even in a unidimensional environment. More clearly, parties are also a cost-sharing technology meaning there are reasons for them to form even in the absence of a benefit from a commitment device. Nevertheless, it is worth analysing further why we find something at least akin to Levy (2004). Given, the similarity of our approach we don't attempt a formal argument, referring the reader to Levy (2004, 2002), but a heuristic argument is provided in Appendix B.

We now compare these results with the empirical distribution of the number of effective parties. Data for 403 post-1945 elections were taken from Kollman, Hicken, Caramani and Backer. (2012), with additional data from Modules 1, 2, and 3 of the Comparative Study of Electoral Systems (2003, 2007, 2013). The combined data are for 59 countries, and the effective number of parties for each election is calculated at the district level as described in Equation 4. Figure 7(c) displays kernel density plots of the number of effective parties in the simulation data for the case of 1-4 dimensions. These are different

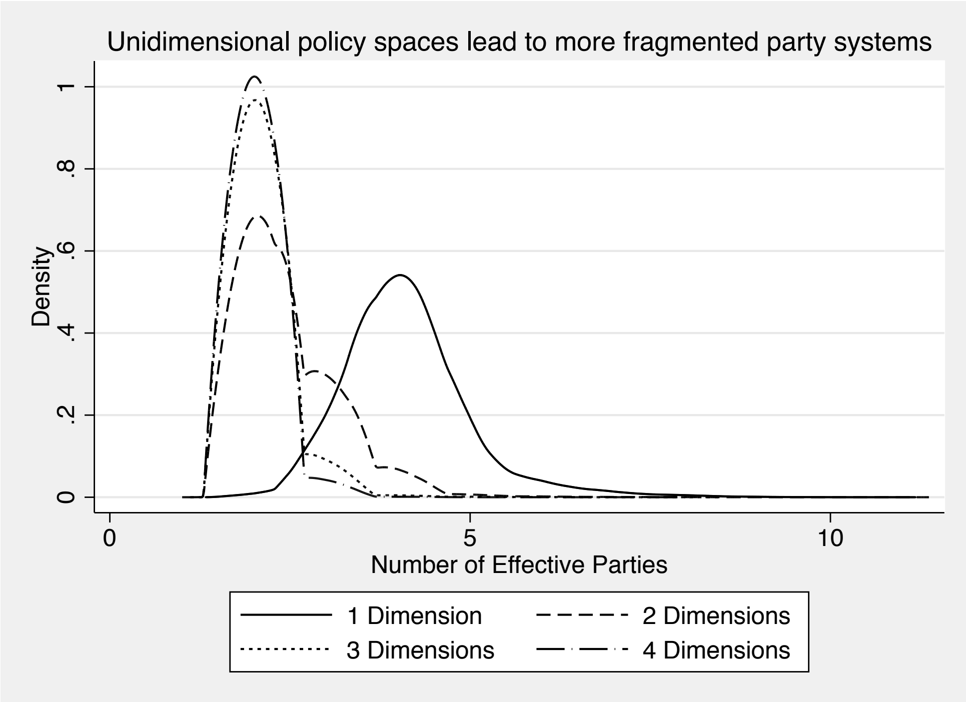
[18] A variety of alternatives to the Laakso-Taagepera (LT) measure have been proposed. One common objection to the LT measure is that it will in general suggest there are several effective parties even when one party has an overall majority and as such only that party is 'effective'. This is less problematic for the purpose here which is to use the effective number of parties as a summary statistic for the overall distribution of party sizes.

to the results in Figure 6 as they include a variety of district sizes weighted to match the empirical data. [19] Results, omitted here for clarity, for between 5 and 7 dimensions are similar to those of 3 and 4 dimensions. This suggests that it is the change from a unidimensional to a multidimensional policy-space that is most important. It is clear that whilst the 1-dimensional case gives rise to too many parties, and the 3 and 4 dimensional cases too few, the 2-dimensional results are extremely similar to those obtained from the empirical data. This is confirmed by Figure 7(d) which overlays the empirical distribution and the 2-dimensional simulation distribution. There is more mass in the right-tail than for the empirical data, but in general it is a very close match. It would be possible to find the combination of results for different numbers of dimensions which minimizes the difference between the moments of the empirical and simulation data. But, the aim is not to emulate exactly the empirical distribution but to argue that the model gives similar results with a minimum of assumptions. It would be inappropriate to argue based on these results that there are in reality often 2 policy dimensions, but the similarity for the case of 2-dimensions is a powerful argument for the importance of considering at least two policy dimensions.

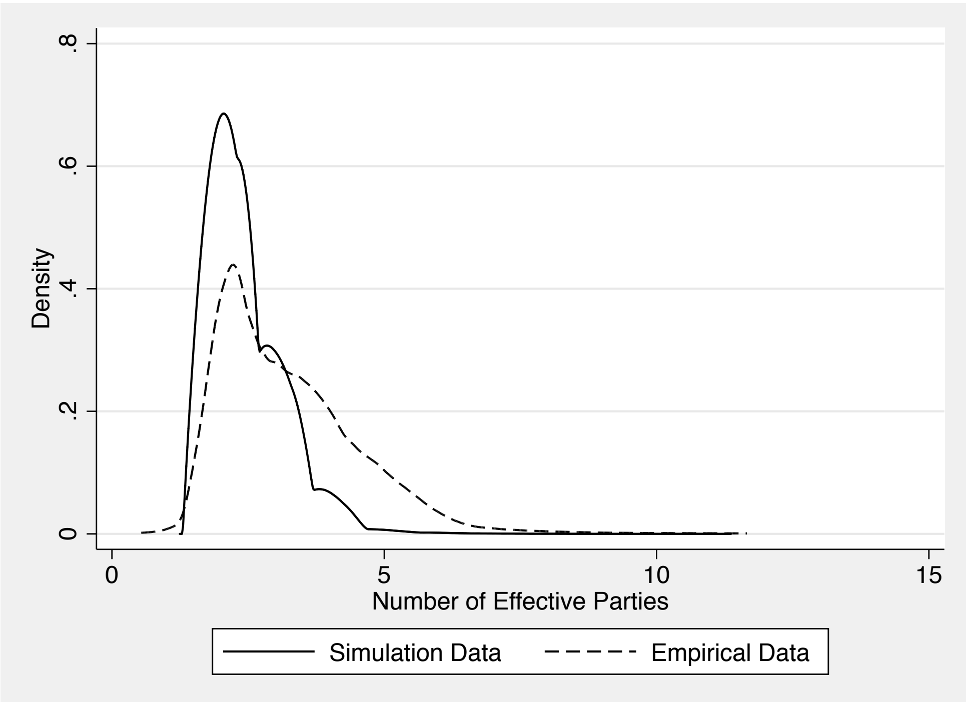
6 Comparative Politics

This section compares the size distributions of parties conditional on average district size predicted by the model with those observed empirically. We find that the relationship between the effective number of parties and district size, is similar to the outcomes observed empirically. Further, we find also find that the model reproduces the empirical relationship between the number of issue-space dimensions and the number of parties. The former relationship has been, and continues to be, the focus of much study, with a large, and growing literature. A pre-eminent discussion of these issues is Lijphart (1999) whilst Gallagher and Mitchell (2005) is an excellent recent survey. Subject to particular attention has been the empirical support and theoretical basis for Duverger's law - *"that plurality elections give rise to a two party system"* (Duverger, 1951).The pre-eminent restatement of

[19] These weights were interpolated by applying a cubic spline to the the empirical data.



(c) Simulation data for Different Dimension Policy Spaces



(d) Comparing 2-Dimensional Simulation and Empirical Data

Figure 7: Comparing Empirical and Simulation Data

this claim is that of Riker (1982), who advocated a statistical interpretation of Duverger's law. Cox (1997) provides a leading summary of the empirical evidence. There has been substantial success in calibrating Riker's interpretation, e.g. Taagepera and Shugart (1993), and more recently Clark and Golder (2006). New theoretical motivation for the

Duvergian hypothesis is, as discussed previously, provided by Morelli (2004). It is this coincidence of theoretical and empirical support that makes it a natural test-bed for the performance of the model developed here. Yet others, such as Dunleavy, Diwakar and Dunleavy (2008), argue that the law has been repeatedly restated in response to ever more violations, and question whether in fact it is '*based on a mistake*'?[20] Whether the empirical data support the 'law' is not a settled matter, and we don't presume to change that here. Our intention is merely to document the consistency between the patterns in actual election data and the results of the model. We argue both are supportive of the weaker, statistical, version of Riker (1982), and argue this is an important validation of the model developed in this paper. We also compare our finding that the number of parties is increasing in the number of policy dimension with the empirical data. We find that although attempts at direct comparison of the effects of dim on N_s are unpromising, the model again is relatively successful at replicating two key empirical relationships.

We begin by visually inspecting the data. Figure 8 plots the average effective number of parties, measured at the constituency level, in 402 post-1945 elections against the (log of) average district magnitude. It suggests that there is a clear positive correlation between the two.

It is clear from Figure 8 that there is a great deal of variation not directly attributable to variation in district magnitude. Given the literature, (cf. Gallagher and Mitchell, 2005), we expect there to be considerable cross-country heterogeneity, but the average district magnitude varies rarely if at all within most countries. Thus, we model country specific characteristics as random effects. Column (1) of Table 2 is the same specification as the line of best fit plotted in Figure 8 but with country-specific random effects. Column (2) introduces a 'Duverger's Law' dummy, equal to one when all seats are selected by single member district elections. The coefficient is negative and large, but only significant at the 10% level. The estimate remains substantially negative, but is measured more precisely when either a linear (Column (3)) or a stochastic (Column (4)) time trend is added.

[20] They emphasise the case of India, where despite a single-member district system there are significantly more than two parties. Importantly, given this empirical debate, Morelli (2004)'s model has the same prediction for the Indian case. In his model if there is sufficient preference heterogeneity (say due to ethnic cleavages) then the Duvergian outcome is reversed.

Column (5) allows for a further non-(log)linearity at large district sizes, by including an analogous variable for a single large district, but this is insignificant.

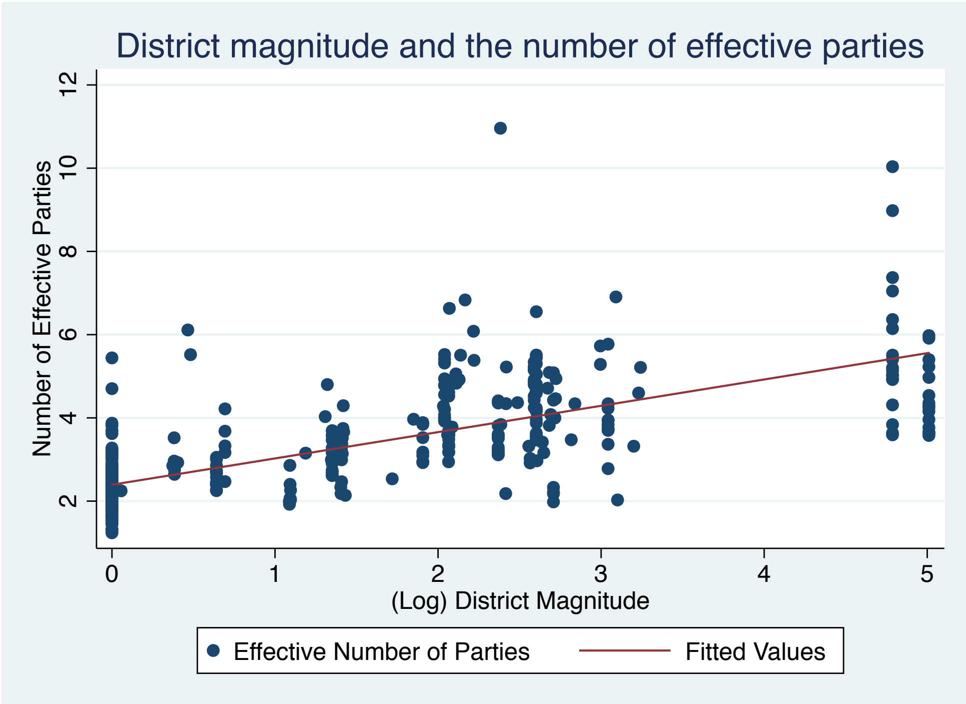


Figure 8: Scatter Plot to show the relationship between District Magnitude and the number of effective parties in 402 elections

We repeat this analysis for the simulation data before considering our other source of variation, the number of policy dimensions. The results of the analysis of the simulation data are reported in Table 1. There are no longer any country or year effects, instead we allow for a fixed effect for each random seed as these are each associated with a given preference distribution and each seed is reused for multiple parameter combinations in order to try and isolate the consequences of *pr* and *dim*. The results are in line with those for the empirical data – higher *pr* leads to more parties, the single member constituency dummy is large and significant, and now the single national district dummy is also.

This close correspondence between the empirical data, which embody any number of national and election idiosyncrasies, and the simulation data, is remarkable. One should be careful in the interpretation of simple models such as the one presented in this paper, but it does suggest that electoral rules are important in determining electoral outcomes, and in a reasonably straightforward manner. Furthermore, the results seem to lend support to a notion of a qualitative distinction between single and multiple-member districts, thus lending support for the probabilistic statement of Duverger's law advanced by Riker (1982).

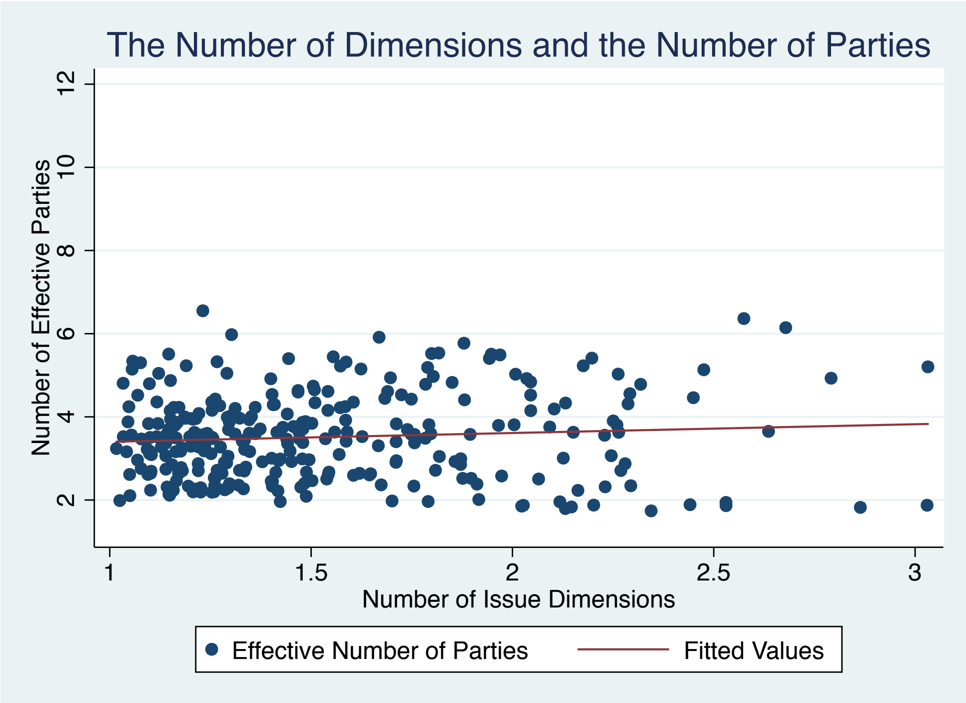


Figure 9: Scatter Plot to show the relationship between the number of policy dimensions and the number of effective parties in 293 elections

We also attempt to compare the relationship between the size distribution of parties and the dimension of the policy-space. Using a novel panel dataset Stoll (2011) studies the relationship between the dimensionality of the policy-space and the number of effective parties. She finds that an increase in the dimensionality leads to an increase in the number of effective parties. This is at odds with the predictions of the model developed here and this discrepancy deserves further explanation. First, if it is important to note that whilst the definition of *pr* for both the empirical and the simulation data are essentially identical, it is harder to have the same confidence about the ideology data. This is natural, in the simulations *dim* is a parameter taking integer values and describing the number of orthogonal, and equally important, policy dimensions. By contrast, the empirical data, are drawn from the manifestos of parties – not the preferences of voters or politicians and thus represent an equilibrium outcome rather than some exogenous input. Thus, despite the sophistication and care of Stoll's approach the resultant data by design measure something quite different to dim.[21] As before, we begin by visual inspection: Compared to Figure 8, Figure 9 reveals little obvious pattern. That Stoll's data measure the dimensionality of each election, rather than a long-run societal average may also be important. Therefore, we also report the time-invariant estimates of Lijphart (1984). The mean is higher, matching better the suggestion of 7(d), but there is only a slight suggestion of a positive correlation.

Columns (7)-(10) of Table 2 contain results describing, first, pooled OLS estimates of the effect of dimensionality on the number of parties, and then random-effect estimates.[22] We note that the pooled estimates find no significant effect, confirming the suggestion of the scatter plot. Unlike the random-effect estimates, where we find, as Stoll does, a positive and significant coefficient on dimension. The interpretation, however, is different. This is a finding that within a given political system more parties are associated with more dimensions, there is no strong evidence for such a relationship across countries, and by extension across institutions. Further, we don't find evidence for an interaction between the number of dimensions and the form of the electoral system, although this may be a consequence of the small available sample. This suggests something else is driving the discrepancy between the within and pooled estimates. In an influential body of work, Taagepera and Grofman (1985), Taagepera and Shugart (1989) and Taagepera (1999) argue for a restatement of Duverger's law that takes into account the variation in the number of policy dimensions across elections and countries. Taagepera and Grofman (1985) and Taagepera and Shugart (1989) argue that Duverger's law is a special case for unidimensional policy-spaces of a more general rule that predicts the number of parties. This rule is that the number of (effective) parties (N_s as defined by equation 4) should be equal to $dim+1$. We thus compute, $\xi = N_s - dim - 1$, these are plotted for the simulation and empirical data in Figure 10.[23] It's clear that whilst there are some differences in the distributions, perhaps unsurprisingly the simulation data are much more dispersed, that

[21] That Stoll's data measure ideology as a continuous variable representing a fractal measure of dimension reveals the difference in the underlying operationalization.

[22] These results are similar to those provided by Stoll (2011) in Table 2 of her paper. We are grateful to her for making her data and further documentation available.

[23] To ease comparison between the two, the simulation data are plotted for $dim = \{1, 2, 3\}$ to correspond to the domain of Stoll's data.

both conform to Taagepera's prediction. More precisely, we are unable to reject the hypotheses that the difference in the two means is zero and that both means are also equal to zero. We argue that the success, or not of the model in replicating, an important feature of the relationship between the number of parties and dimension in the empirical data is suggestive that despite inevitable differences in definition and measurement, the model is largely a success in understanding the consequences of variation in preferences for democratic outcomes.[24]

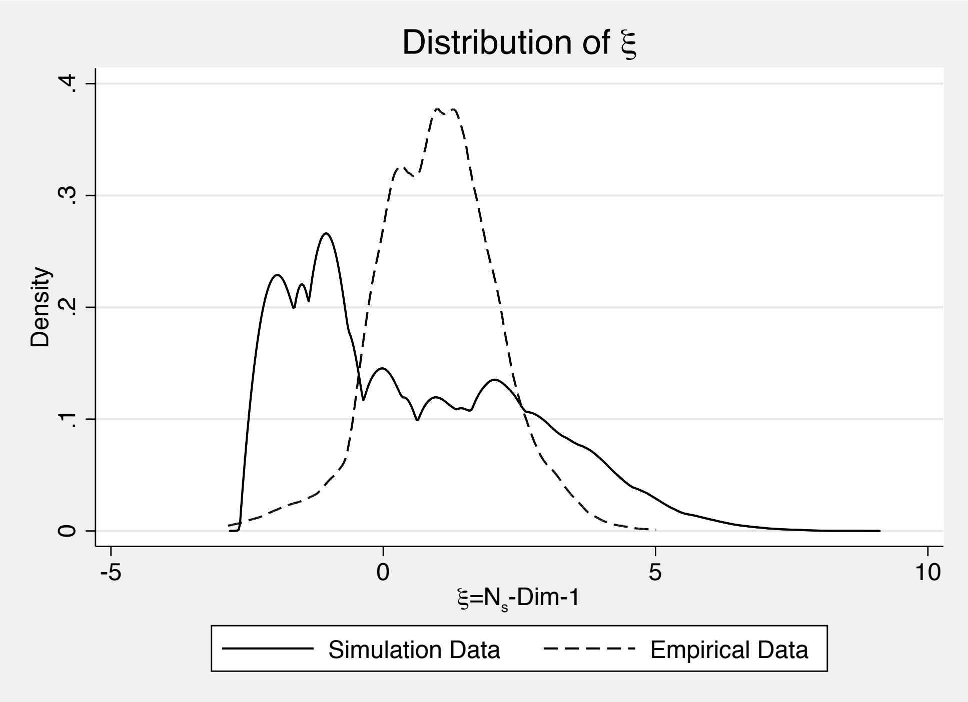


Figure 10: Distribution of ξ in the simulation and empirical data. ξ is the deviation from the hypothesised relationship that the Number of effective parties equals the number of policy dimensions minus one.

Taagepera (1999) sought to combine the insights of this work on the effects of ideology, with that studying Duverger's law to model the relationship between district magnitude, the number of policy dimensions, and the number of effective parties. He proposed, in the notation of our model, the following relationship:

$$N_s = dim^{0.6} pr^{0.15} + 1 \tag{5}$$

The multiplicative form follows, from the argument of Ordeshook and Shvetsova (1994)

[24] Using Lijphart's data produces a much closer correspondence in the two graphs.

for the case of ethnic heterogeneity and Neto and Cox (1997) for issues (or 'cleavages') more generally. They argue, that the consequences of issues and electoral rules are not independent. For example, a society with many cleavages, is more likely to give rise to many parties when there are multiple-member districts that facilitate this, and similarly such districts will be less likely to lead to more parties when there are fewer cleavages. Taagepera (1999) takes this insight and develops the parametrization above. As above, after subtracting the righthandside terms, we apply 5 to both the simulation and the empirical data. We plot the resulting series in Figure 11, this time referring to the prediction error as ϕ. The similarity between the empirical and simulation distributions of ϕ is obvious, the model reproduces an almost identical distribution of deviations from the parametrization proposed by Taagepera (1999).[25] To be able to match the joint impact of preferences and institutions observed empirically suggests the usefulness of the approach advanced in this paper. The next section uses individual level data to analyze the welfare implications of variation in electoral rules.

7 Welfare Comparisons

A fundamental reason to be interested in party systems is because we believe they might lead to differences in welfare. Given that we calculate each individual's utility as part of solving the model it is relatively straightforward to consider whether overall welfare varies with either *pr* or *dim*. That is we directly compare aggregate welfare across different political systems based on individual utilities to analyze the different mechanisms responsible for the observed variation. We find that an increase in the number of representatives per district is associated with a reduction in welfare, and that this is despite an associated increase in welfare due to an increased number of political parties. Our results are relevant in both the context of the large literature in Comparative Politics that seeks to compare the outcome of different electoral systems and in the context of the social-choice literature. In particular our approach is well suited to analysing the

[25]Given that we use a different dataset an alternative parametrization might match the empirical data better. However, we prefer to use Taagepera's in the interests of both parsimony and as it represents an objective test.

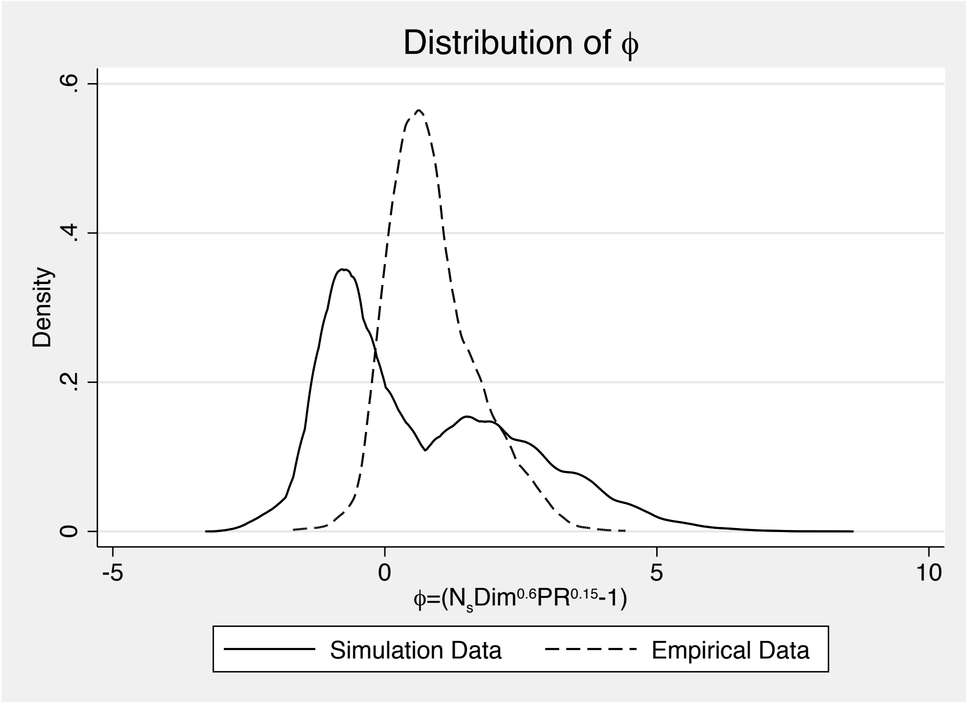


Figure 11: Empirical and Simulation data distributions of $\phi = dim^{0.6}pr^{0.15} + 1 - N_s$ – the deviation from the hypothesis that the number of parties is jointly increasing in dim and pr.

consequences of discrete differences between systems, in this sense it is closer in spirit to Persson, Roland and Tabellini (2000) and Persson and Tabellini (1999, 2000, 2004) than to Lijphart (1999). However, it remains the case following Arrow (1950) that we are unable to make definite statements of the superiority of one system or another from a welfare perspective. Secondly, all of our results are contingent on our choice of utility function and on the assumed distributions of preferences. Given that here utility is a continuous variable, and we consider finite populations, we restrict ourselves to a simple symmetric utilitarian social-welfare function.[26]

We first consider whether welfare varies with the form of democracy (pr). Column 1 of Table 3 reports simple bivariate regression results. We find that an additional representative in a district is associated with a reduction in welfare, by around 0.11 (24%). This effect is large, and statistically significant, but confounded with other consequences of changes in the number of representatives per district: Previously, we found that multi-member districts were associated with more political parties. However, partialling out the

[26]The key results below are robust to the use of a minimax social-welfare function.

effect of additional parties reveals that district size has an effect in excess of that mediated by the number of parties. Column (2) reports that the coefficient on pr now suggests a reduction in welfare of just over 0.04 (8.6%) whilst each additional party increases welfare by around 0.11. Column (3) seeks to clarify the mechanism at work by also including dim, columns (4) and (5) include the interactions $pr \times dim$ and $numparties \times dim$. The results imply that even though the interaction $pr \times dim$ is positive, the effects of pr is negative for all parameter values that we consider. In contrast the effect of the number of parties whilst positive for values of $dim \leq 2$ is negative in higher dimensional spaces. Column (6) repeats the exercise but calculates the utility of the median voter, or for $dim > 1$ the utility of the voter closest to the median preference across all dimensions.[27] The only notable difference is that the R^2 of this regression is about 0.92 versus around 0.74 for the specification in Column (5). This suggests that a considerable proportion of the unexplained variation may be attributed to the variation in utilities realized by voters with different preferences for a given outcome. The results imply that there is an effect of pr on welfare beyond that attributable to the number of parties, that is that the largest party is (on average) further from the mean voter in multimember districts. This result is surprising, multi-member districts are traditionally argued to improve representativeness, but here we find that whilst the increased number of parties associated with multi-member districts does improve welfare (and here by implication representativeness), other properties of the resultant distribution of parties mean that overall welfare is reduced. Moreover, the improvement in welfare associated with more political parties is reversed in higher-dimensional policy-spaces. We explore this result further below. Of course, what our model ignores is post-election coalition formation and it may well be, as argued by Lijphart (1999), that this leads to what he describes as *'kinder, gentler, outcomes'*. Equally, it may be other variations in the form of democracy, not modelled here, that lead to these outcomes. Nevertheless, the result suggests that

[27]Given the median voter will (in general) fail to exist in multiple dimensions, we define the pseudo-median voter in district j as

$$voter_{mj} s.t \operatorname{argmin} m \sum_{k=1}^{K} (\tilde{a_k} - a_{mk})^2$$

where a $\tilde{a_k}$ is the preference of the median voter on dimension k.

there are important consequences of variation in electoral system beyond the impact on the size-distribution of political parties. Further, a key advantage of the approach taken in this paper is that the welfare claims advanced in this section emerge directly from the model and require no ancillary assumptions as to what constitutes 'good' or 'bad' political outcomes.

Another implication of the results in Table 3 is that welfare is decreasing in the dimensionality of the policy-space. Although more dimensions are associated with fewer parties, the conditional effect of dimension on welfare is almost identical. This result is unsurprising – even though utilities are normalized to keep the total size of the issue space constant, such that $U^*_{ij} = U_{ij}/\sqrt{N}$, given that 1 is quadratic then the average utility loss associated with an additional dimension is $E[w_k^2]$. Given preferences are approximately normally distributed, then the average welfare loss for each additional dimension is given by $E[N(\mu, \sigma)^2] = E[N(0.5, 0.3)^2] \approx 0.3$. We thus repeat the analysis in Table 3 but using $U^1_{ij} = w_1 - a_1$. The coefficient is now several orders of magnitude smaller, but still suggests that there is an impact of dimension beyond that of its effect on the number of parties. One further possibility is that the welfare consequences of dimension vary with pr, but whilst an interaction term, as reported in Column (4) is positive the overall effect remains negative (and significant) for all parameter combinations. This specification also includes binary variables for each level of the number of parties variable. We are thus forced to consider an explanation reliant on differences in the equilibrium configuration of parties as dimension increases, such that the average policy is further away from the average voter. Why this might be will be the subject of future research.

8 Conclusion

This paper has presented a new approach to the analysis of party systems. Developing and simulating a citizen-candidate model with endogenous party formation with which to study the effects of variation in preferences and institutions. The results suggest that not only does such a model give rise to qualitatively realistic outcomes, but also reproduces key features of the empirical data. Section 5 provides new evidence of the importance of multiple dimensions for party formation. The distribution of the empirical data were found to be similar to the case of two policy dimensions. Section 6 showed that the model reproduces the key empirical relationships between district size and the number of parties, and in particular provided new support for Duverger's law. Furthermore, evidence was provided that the model reproduces the interaction between institutions and preferences as observed in the data. Finally, Section 7 suggested that multi-member districts are associated with lower welfare, and that higher dimension policy-spaces increase welfare by engendering more parties, but conditional on their number, reduce welfare. Crucially, as argued in Sections 2 and 3 all of these results may be regarded as reflecting equilibrium behavior of strategic agents. The approach advocated by this paper may be used productively in the future to study both particular historical events in more detail, for example electoral reform, and also more complicated representations of institutions, for example party lists or mixed voting systems, or preferences, for example extremism.

References

Arrow, Kenneth J, "A difficulty in the concept of social welfare," *The Journal of Political Economy*, 1950, *58* (4), 328–346.

Baylie, Dan and Guy Nason, "Cohesion of Major Political Parties," *British Politics*, 2008, *3*, 390–417.

Bendor, Jonathan, Daniel Diermeier, David A Siegel, and Michael M Ting, *A behavioral theory of elections*, Princeton University Press, 2011.

Besley, Timothy and Stephen Coate, "An Economic Model of Representative Democracy," *The Quarterly Journal of Economics*, 1997, *112* (1), 85–114.

Clark, William Roberts and Matt Golder, "Rehabilitating duvergers theory testing the mechanical and strategic modifying effects of electoral laws," *Comparative Political Studies*, 2006, *39* (6), 679–708.

Cox, Gary W, *Making votes count: strategic coordination in the world's electoral systems*, Vol. 7, Cambridge Univ Press, 1997.

Degan, Arianna and Antonio Merlo, "Do voters vote ideologically?," *Journal of Economic Theory*, 2009, *144* (5), 1868 – 1894.

Dhillon, Amrita, "Political Parties and Coalition Formation," in Gabrielle Demange and Myrna Wooders, eds., *Group Formation in Economics: Networks, Clubs, and Coalitions*, Cambridge University Press, 2005, chapter 9.

Downs, Anthony, *An Economic Theory of Democracy*, New York: Harper & Row, 1957.

Dunleavy, Patrick, Rekha Diwakar, and Christopher Dunleavy, "Is Duverger's Law based on a mistake?," 2008. Political Studies Association Annual Conference Paper.

Dutta, Bhaskar, Matthew O. Jackson, and Michel Le Breton, "Strategic Candidacy and Voting Procedures," *Econometrica*, 2001, *69* (4), 1013–1038.

Duverger, Maurice, *Political Parties: Their Organization and Activity in the Modern State*, Wiley, 1951.

Eguia, Jon X., "On the spatial representation of preference profiles," *Economic Theory*, 2013, *52* (1), 103–128.

Gallagher, Michael and Paul Mitchell, *The Politics of Electoral Systems*, 1st ed., Oxford University Press, 2005.

Golub, Gene H. and Charles F. Van Loan, *Matrix Computations*, 3rd ed., John Hopkins University Press, 1996.

Hart, Sergiu and Mordecai Kurz, "Endogenous formation of coalitions," *Econometrica: Journal of the Econometric Society*, 1983, pp. 1047–1064.

Humphreys, Macartan and Michael Laver, "Spatial models, cognitive metrics, and majority rule equilibria," *British Journal of Political Science*, 2010, *40* (01), 11–30.

Kollman, Ken, Allen Hicken, Daniele Caramani, and David Backer., "Constituency-Level Elections Archive (CLEA; www.electiondataarchive.org)," Ann Arbor, MI: University of Michigan, Center for Political Studies [producer and distributor] December 2012.

_ , John H. Miller, and Scott E. Page, "Adaptive Parties in Spatial Elections," *The American Political Science Review*, 1992, *86* (4), 929–937.

_ , _ , and _ , "Political Institutions and Sorting in a Tiebout Model," *The American Economic Review*, 1997, *87* (5), 977–992.

Laakso, M. and Rein Taagepera, "Effective Number of Parties: A Measure with Application to West Europe," *Comparative Political Studies*, 1979, *12*, 3–27.

Laver, Michael, "Policy and the Dynamics of Political Competition," *The American Political Science Review*, 2005, *99* (2), 263–281.

_ and Ernest Sergenti, *Party Competition: An Agent-Based Model: An Agent-Based Model*, Princeton University Press, 2011.

Levy, Gilat, "Endogenous Parties: Cooperative and Non-cooperative Analysis," mimeo, LSE and TAU 2002.

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

Lijphart, Arend, "Democracies: Patterns of majoritarian and consensus government in twenty-one countries," 1984.

_ , *Patterns of Democracy: Government Forms and Performance in Thirty-six Countries*, paperback ed., New Haven, London: Yale University Press, 1999.

Lloyd, Stuart P., "Least Squares Quantitization in PCM," *IEEE Transactions on Information Theory*, 1982, *28* (2), 129–137.

Merlo, Antonio M., "Whither Political Economy? Theories, Facts and Issues," in Richard Blundell, Whitney Newey, and Torsten Persson, eds., *Advances in Economics and Econometrics: Theory and Applications: Ninth World Congress of the Econometric Society*, Vol. 1, Cambridge: Cambridge University Press, 2006, pp. 381–421.

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

Neto, Octavio Amorim and Gary W Cox, "Electoral institutions, cleavage structures, and the number of parties," *American Journal of Political Science*, 1997, pp. 149–174.

of Electoral Systems, University of Michigan Comparative Study, "The Comparative Study of Electoral Systems (www.cses.org). CSES Module 1," www.cses.org August 2003.

_ , "The Comparative Study of Electoral Systems (www.cses.org). CSES Module 2," www.cses.org June 2007.

_ , "The Comparative Study of Electoral Systems (www.cses.org). CSES Module 3," www.cses.org March 2013.

Ordeshook, Peter C and Olga V Shvetsova, "Ethnic heterogeneity, district magnitude, and the number of parties," *American journal of political science*, 1994, pp. 100–123.

Osborne, Martin J. and Al Slivinski, "A Model of Political Competition with Citizen-Candidates," *Quarterly Journal of Economics*, Febuary 1996, *111* (1), 65–96.

_ **and Rabee Tourky**, "Party Formation in Single-Issue Politics," *Journal of the European Economic Association*, 09 2008, *6* (5), 974–1005.

Ostrom, Elinor, "An Agenda for the Study of Institutions," *Public Choice*, 1986, *48* (1), pp. 3–25.

Persson, T., G. Roland, and G. Tabellini, "Comparative politics and public finance," *Journal of Political Economy*, 2000, *108* (6), 1121–1161. 50.

Persson, Torsten and Guido Tabellini, "The size and scope of government:: Comparative politics with rational politicians," *European Economic Review*, 1999, *43* (4), 699–735.

_ **and** _ , *Political Economics: Explaining Economic Policy* Zeuthen Lecture Books, 1st ed., Cambridge, MA: The MIT Press, 2000.

_ **and** _ , "Constitutional rules and fiscal policy outcomes," *American Economic Review*, 2004, pp. 25–45.

Plott, Charles R., "The Application of Laboratory Experimental Methods to Public Choice," Working Papers 223, California Institute of Technology, Division of the Humanities and Social Sciences 1979.

Ray, Debraj and Rajiv Vohra, "Equilibrium binding agreements," *Journal of Economic Theory*, 1997, *73* (1), 30–78.

Riker, William H., "The Two-Party System and Duverger's Law: An Essay on the History of Political Science," *The American Political Science Review*, 1982, *76* (4), pp. 753–766.

Roemer, John E., "The Democratic Political Economy of Progressive Income Taxation," *Econometrica*, 1999, *67* (1), 1–19.

Rustichini, Aldo, "Optimal Properties of Stimulus-Response Learning Models.," *Games and Economic Behavior*, October 1999, *29* (1-2), 244–273.

Snyder, James M. and Michael M. Ting, "An informational rationale for political parties," *American Journal of Political Science*, 2002, *46* (1), 90–110.

Stoll, Heather, "Dimensionality and the number of parties in legislative elections," *Party Politics*, 2011, *17* (3), 405–429.

Taagepera, Rein, "The Number of Parties As a Function of Heterogeneity and Electoral System," *Comparative Political Studies*, 1999, *32* (5), 531–548.

_ **and Bernard Grofman**, "Rethinking Duverger's Law: Predicting the Effective Number of Parties in Plurality and PR Systems–Parties Minus Issues Equals One*," *European Journal of Political Research*, 1985, *13* (4), 341–352.

_ **and Matthew Soberg Shugart**, *Seats and votes: The effects and determinants of electoral systems*, Yale University Press New Haven, 1989.

_ **and** _ , "Predicting the Number of Parties: A Quantitative Model of Duverger's Mechanical Effect," *The American Political Science Review*, 1993, *87* (2), pp. 455–464.

Tullock, Gordon and Colin D. Campbell, "Computer Simulation of a Small Voting System," *The Economic Journal*, March 1970, *80* (317), 97–104.

Table 1: Analysis of the Simulation Data

	(1) effparty	(2) effparty	(3) effparty	(4) effparty	(5) effparty	(6) effparty
lpr	0.400***	0.305***	0.364***		0.264***	0.631***
	(0.010)	(0.012)	(0.012)		(0.009)	(0.011)
single		-0.377***	-0.311***		-0.400***	
		(0.018)	(0.017)		(0.016)	
large			-0.768***		-0.649***	
			(0.058)		(0.053)	
dim				-0.895***	-0.874***	-0.788***
				(0.010)	(0.009)	(0.006)
prdim						-0.015***
						(0.000)
_cons	2.739***	2.938***	2.872***	5.378***	5.146***	4.629***
	(0.010)	(0.017)	(0.016)	(0.038)	(0.033)	(0.021)
N	16455	16455	16455	16455	16455	16455

Standard errors, clustered by random number seed, in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Analysis of the Empirical Data

	(1) ENP_wght	(2) ENP_wght	(3) ENP_wght	(4) ENP_wght	(5) ENP_wght	(6) ENP_wght	(7) ENP_wght	(8) ENP_wght	(9) ENP_wght
lmag	0.652^{***}	0.517^{***}	0.442^{***}	0.420^{***}	0.422^{***}	0.375		0.404^{***}	0.343^{***}
	(0.086)	(0.066)	(0.147)	(0.147)	(0.138)	(0.240)		(0.084)	(0.132)
single		-0.468^{***}	-0.633^{*}	-0.753^{*}	-0.805^{**}	-0.873^{*}		-0.453^{**}	
		(0.158)	(0.356)	(0.385)	(0.360)	(0.491)		(0.202)	
large						0.305			
						(0.969)			
rawdim							0.211	0.544^{***}	0.354^{*}
							(0.185)	(0.170)	(0.214)
dim_lmag									0.073
									(0.092)
N	402	402	402	402	402	402	293	210	210

Standard errors, clustered by country, in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Welfare Analysis

	(1) utility	(2) utility	(3) utility	(4) utility	(5) utility	(6) medvot	(7) altutility	(8) altutility
pr	-0.010^{***}	-0.034^{***}	-0.010^{***}	-0.032^{***}	-0.010^{***}	-0.003^{***}	0.005^{***}	0.024^{***}
	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
dim			-0.070^{***}	-0.086^{***}	-0.069^{***}	-0.000	-0.006^{***}	0.006^{***}
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
prdim				0.009^{***}				-0.007^{***}
				(0.000)				(0.000)
numparties_constdim					-0.000^{***}	-0.000^{***}		
					(0.000)	(0.000)		
N	132252000	132252000	132252000	132252000	132252000	132252000	132252000	132252000

Standard errors, clustered by run, in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Party Formation and Competition - Appendix

Daniel Ladley

Department of Economics
University of Leicester*

James Rockey

Department of Economics
University of Leicester

September 2013

A Solving the Model

A.1 The choice of Norm

The choice of metric, and therefore the class of preferences considered, may be thought to be important given this paper's focus on multidimensional policy-spaces. Two obvious alternatives are the Euclidean (l_2) and Manhattan (l_1) norms. The first has the advantage of corresponding to our standard geometric intuitions, with distances given by the Pythagorean theorem. The second has been advocated in recent work, including Eguia (2013a), Eguia (2013b), and Humphreys and Laver (2010) and has the alternative intuition that the total disagreement (distance) across all issues is the sum of all of the individual disagreements (distances). Thus, if two individuals are say, 1 unit of policy-space apart on each of two issues the total disagreement is 2 not $\sqrt{2}$. This is appealing, if as here the assumed preference dimensions are best thought of as representing a series of orthogonal philosophical or attitudinal fundamentals, rather than specific (potentially correlated) policy choices.[1]. The less-obvious but perhaps more important difference between the two norms is that if l_1 preferences are assumed then the marginal utility of a a deviation

*Email: Ladley-dl110@le.ac.uk, Rockey-jcr12@le.ac.uk

[1]For example, the first may represent fiscal-conservatism, the second social-conservatism, etc. Alternatively, these may be thought of in statistical terms as the principal components of a set of substantive policy issues

in any one dimension is independent of the difference on all other dimensions. This seems overly restrictive, and might rule out a variety of interesting trade-offs, we therefore prefer the Euclidean Norm as it is standard in the previous literature. Re-running the model with the l_1 suggests that the choice is unimportant for our results.

A.2 A single district election

The model is solved iteratively. Citizens initially treat the decision to stand as a coin toss, and then through experimentation learn their pay-off from standing or not. This pay-off is conditional on all other citizens' decisions to stand or not. To make this process as clear as possible, we first consider a sandbox model with one constituency without coalitions. Let $T = [0, \ldots, t, \ldots, T]$ index time periods. At $T = 0$ preferences are generated from a given distribution based on a random number seed, and the other model parameters are set. Then events proceed as follows:

1. All individuals within each constituency simultaneously declare whether they will stand for office. This results in the set C of candidates. They do so to maximize, as in the main text, the following set of payoffs:

$$U^j(W) = \begin{cases} 1 - \frac{|W - A_j|}{\sqrt{N}} - \kappa + \gamma & \text{if she is elected} \\ 1 - \frac{|W - A_j|}{\sqrt{N}} - \kappa & \text{if she is not elected} \\ 1 - \frac{|W - A_j|}{\sqrt{(N)}} & \text{if she does not stand} \end{cases} \tag{1}$$

An individual's payoff from standing is conditional on the standing decisions of all other citizens. Solving this problem directly is likely difficult, and instead the set of optimal strategies is obtained by simulating individuals as learning their optimal strategy through experimentation. In each election (period) candidates decide to run with probability P_{jt}^{stand} which is based on the relative utility derived from standing or not in previous periods. Specifically,

$$(2) \qquad P_{jt}^{stand} = \frac{U_{jt}^{Run}}{U_{jt}^{NoRun} + U_{jt}^{Run}}$$

Where U_{jt}^{Run} and U_{jt}^{NoRun} are the total of individual j's utility from running or not running in all previous periods. After each election each individual, given an implemented policy W, calculates $U_{j,t+1}^{Run}$ and $U_{j,t+1}^{NoRun}$ as follows:

$$(3) \qquad U_{j,t+1}^{Run} = \beta U_{j,t}^{Run} + \mathbf{1}(stood_t)U^j(W)$$

$$(4) \qquad U_{j,t+1}^{NoRun} = \beta U_{j,t}^{NoRun} + (1 - \mathbf{1}(stood_t))U^j(W)$$

As discussed in the main text, Rustichini (1999) shows that when counterfactuals are not observed as is the case here, that a linear learning rule such as equation 3 will lead citizens to converge to the optimal action. However, $1-\beta$ can be regarded as the rate at which citizens discount (forget) their utilities from previous choices. Only if $\beta = 1$, when players treat all previous outcomes equally, will Rusticini's result hold. However, this has to be balanced with a practical requirement that the algorithm eventually converges. Experimentation suggests that $\beta = 0.99$ is a good choice, the results do not change with a higher-value, while it gives adequate convergence speeds. The other parameter governing the learning rule is the initial probability of standing:

Where:[2]

$$Uj,0^{Run} = U_{j,0}^{NoRun} = 1 \Leftrightarrow P_{j,0}^{stand} = 0.5.$$

2. Every individual $j \in J$ simultaneously, sincerely, votes. That is, they vote for the candidate who's ideal point is closest to their own:

[2]The values of $U_{j,0}^{NoRun}$ and $U_{j,0}^{NoRun}$ have a limited effect on model behavior, larger values can dramatically increase the time until convergence and setting either value to be less than or equal to zero can cause obvious convergence problem.

$$\arg\min_{k} |(A_k - A_j)| \text{ where } k \in C \tag{5}$$

Votes for each candidate are counted and the m individuals from each district with the highest number of votes are elected to office. If $|C^d| < m$ all members of C^d are elected, however, all non-standing members of d receive $-\kappa - 1$ utility for this election. This large negative utility is of greater magnitude than the lowest utility an individual may receive if they stand for election and so ensures that once the model has converged there will be at least m individuals standing for election in each district.[3]

3. Payoffs are realized according to 1. Citizens update their probability of standing according to 2 and 3.

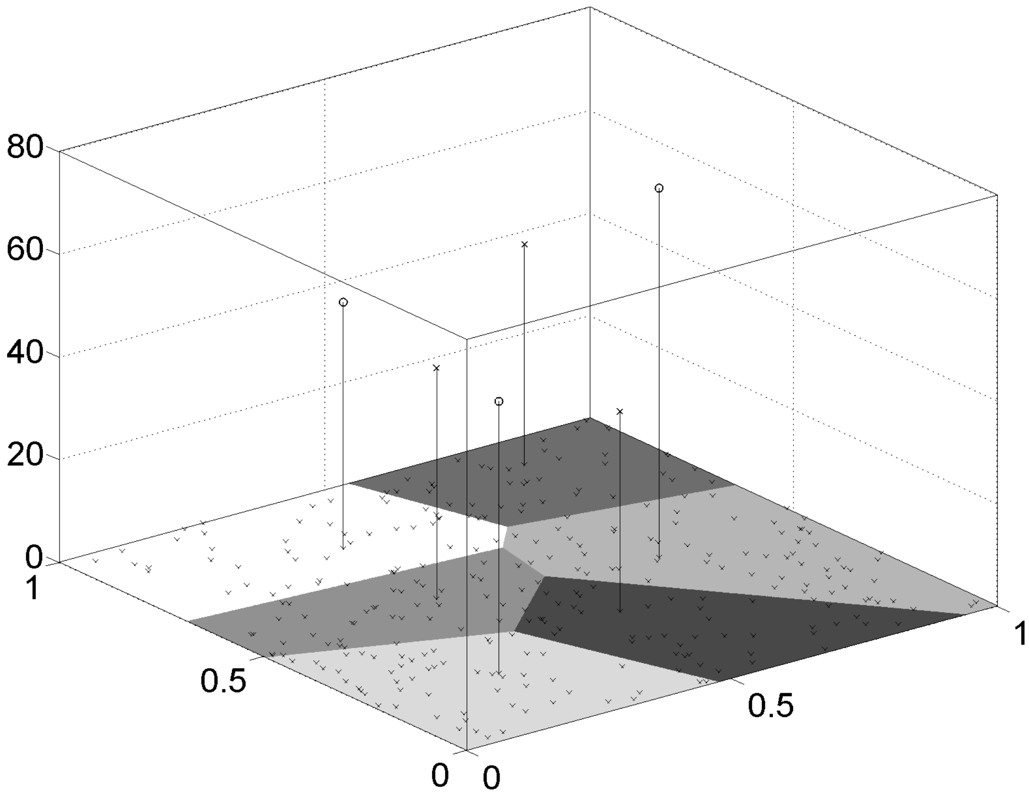
4. The process is repeated for a large number of iterations until an equilibrium has been reached. This is defined as where $P^{stand}_{jt} = \{0, 1\} \forall j$, that is every citizen either stands or not with certainty. This rules out outcomes analogous to mixed strategy equilibria in which some citizens stand with probability $0 < P^{stand}_{jt} < 1$ but in which the distribution of standing probabilities across all citizens has become stable. Such equilibria are discarded along with those cases for which an equilibrium fails to exist.[4]

To illustrate what such an equilibrium might look like it is instructive to consider an example. Figure A.2 is one realization for the case of single district from which three candidates will be elected and where there is a two-dimensional policy space. The individual x's on the xy-plane represent the preferred policy of individual citizens who have chosen not to stand. The stalks with elevated x's are candidates who in equilibrium

[3] This payoff is analogous to the negative infinity payoff received when insufficient candidates stand for election in the Osborne and Slivinski (1996) model.

[4] We define these as when the simulation has run for 10^9 iterations without showing signs of convergence. These account for a relatively small number of seeds, and although is more common for higher values of pr and dim, remains fewer than 1 in $10,000$

stand for election but don't win. This is in common with the OS and BC where candidates run in order to alter the identity of the winner, as this change in winner offsets the cost of standing. The shaded areas on the xy-plane are a Voronoi tessellation,as employed in the context of voting by Degan and Merlo (2009) for a general treatment see Okabe, Boots, Sugihara and Chiu (2009). This describes the areas of the preference space corresponding to each candidates support. That is, they are the half-spaces in which all citizens vote for a given candidate. For a set of party platforms, $\mu = \mu_{rk}$, for parties $r = (1, \ldots, N)$ in dimensions $k = 1, \ldots, K$ then the regions P_r are such that for every voter with blisspoint $a_{jk} \in P_r, d(a, \mu_r) \leq d(a, \mu_{s \neq r}) \forall \mu_s \in \mu$, where here $d(\cdot, \cdot)$ is the Euclidean norm. These are depicted by the shading of the policy space (with a Voronoi tesselation) Speculatively, the three non-winning candidates can be seen to be anchoring the (identity of the) winning candidates away from the center which may otherwise have been a (Downsian) equilibrium.



A.3 Party Membership

We seek to find a stable partition of politicians given that each politician wishes to maximize:

$$V_r^j = \frac{\#r}{\sqrt{\sum_{k=1}^{N}(w_{ik} - \mu_k)^2 + \eta^2}} \tag{6}$$

The process of coalition formation proceeds as follows. Initially each newly elected representative starts a new coalition of which they are, at this point, the only member. All returning representatives remain in their previous coalition, whether or not all previous members have been re-elected. Once, all representatives belong to a coalition (possibly with a total membership of 1)[5], candidates assess whether their current coalition best represents their interests.

We implement this computationally in two stages. The algorithm first identifies groups or individuals within each coalition most likely to wish to secede and then tests whether either they (or the rest of their party) would have higher average utility after a split. The second step of the algorithm tests whether random pairs of coalitions would improve their average utilities by merging. Through repeated applications of this algorithm within each period $t = 1, \ldots, t, \ldots, T$ this algorithm will identify an allocation of party memberships that is 'bi-Core' stable, as defined by Ray and Vohra (1997).

The above process occurs after each election, each coalition (in random order) first tests whether it would be beneficial if it splits and then tests whether it would be beneficial if it merged with other randomly chosen coalitions. Once, the membership of the coalitions has been established it is assumed that the preferences of the coalition with the most representatives are implemented. This is an abstraction, for instance it is not necessarily the case, as in observed democracies, that the largest coalition contains a majority of representatives. However, the focus here is on the electoral process and not on the process of government policy formation. All individuals in all districts therefore receive payoffs

[5] Coalitions with no-members are assumed to no-longer exist.

based on the implemented policy of the largest party.

The composition of coalitions changes through a process of splitting and merging.[6] These processes identify whether there are subsets of coalitions that would be better off as separate coalitions or whether there exists pairs of parties which would be better off if they merged. As such it is a coalition-stability concept.

In order to conduct the splitting analysis principle groupings are found within each party using the k-means algorithm as first proposed by Lloyd (1982) and as interpreted by Hartigan and Wong (1979). This algorithm is widely used to identify clusters in multi-dimensional data. In essence it searches for the allocation of observations to clusters and the means of those clusters that minimizes the total sum of the squared distances between cluster midpoints and the points in each cluster, across all clusters. Here, we employ it to partition each coalitions into two groups who each consider whether it is in their interest to leave the coalition. In particular, the j candidates are partitioned into z sets (here $z = 2$). This collection of sets $G = \{G_1, .., G_z\}$ is chosen so to minimize the total within group variance, across all groups. That is:

$$\underset{G}{\operatorname{argmin}} \sum_{i=1}^{z} \sum_{A_j \in G_i} |(A_j - \mu_i)|^2 \tag{7}$$

The algorithm to do this proceeds in the following steps:

1. Initially two 'centers' P_1 and P_2 are chosen at random within the policy space.
2. Each member of the coalition identifies which centre they are closest to producing two groups G_1 and G_2
3. Set P_1 equal to the mean of the ideal points of G_1 and similarly for P_2 and G_2
4. Repeat from 2 until the centers no longer change.

This algorithm is not deterministic, it is dependent on the initially chosen centres and may find different clusters each time it runs. This is advantageous for this model as it

[6] We considered an additional process whereby individuals could unilaterally change coalition if under the above metric it was beneficial to do so. It was found that this did not effect the distribution of results.

allows a more thorough testing of the stability of each coalition as different groups consider seceding. Once the groups have been identified each group must determine whether to break away. Their decision is based on the satisfaction of the individuals in the cluster with their continued membership. The average utility of the members of each group is calculated as a combined party and as separate coalitions. If the average utility of either group is higher after seceding then the coalition splits. The decision to secede is a unilateral one, a group does not need permission to leave a coalition.

Similarly each coalition considers if it would be better off merging with another party chosen at random. If the average utility of the members of both parties are greater as a combined unit than as separate groupings the two coalitions merge into one. In this case it is necessary that both groupings increase their utility for the merger to occur.

When both splitting and merging the average utility of members of the groups are employed in decision making. Consequently there may be one or more members of each group which disagree with the decision. In the long term, however, this dissatisfaction does not persist, the potential for coalitions to further split and merge, or the citizen potentially no-longer standing, ensures that eventually each individual is happy with their final position.

It is worth emphasizing that the two elements below, 'citizens vote' and 'candidates vote' occur together and the results are amalgamated in order to determine those in office. Flowchart 2 shows the inter-coalition procedures, unlike Flowchart 1 this is not done from an individual perspective, rather it considers a top-down view of the model.

Flowchart 1 shows the decisions made by citizens as the above model proceeds, whilst to ease readability Flowchart 2 expands on the details of the party dynamics box of Flowchart 1. As described above the model commences with the setting of the preference distributions and continues until convergence. Each election starts with citizens declaring their candidacy and finishes when individuals calculate their payoffs. If the model has not reached an equilibrium, e.g. standing probabilities aren't all zero or one, citizens learn according to the rules described above and another election is called.

A.4 Flowcharts

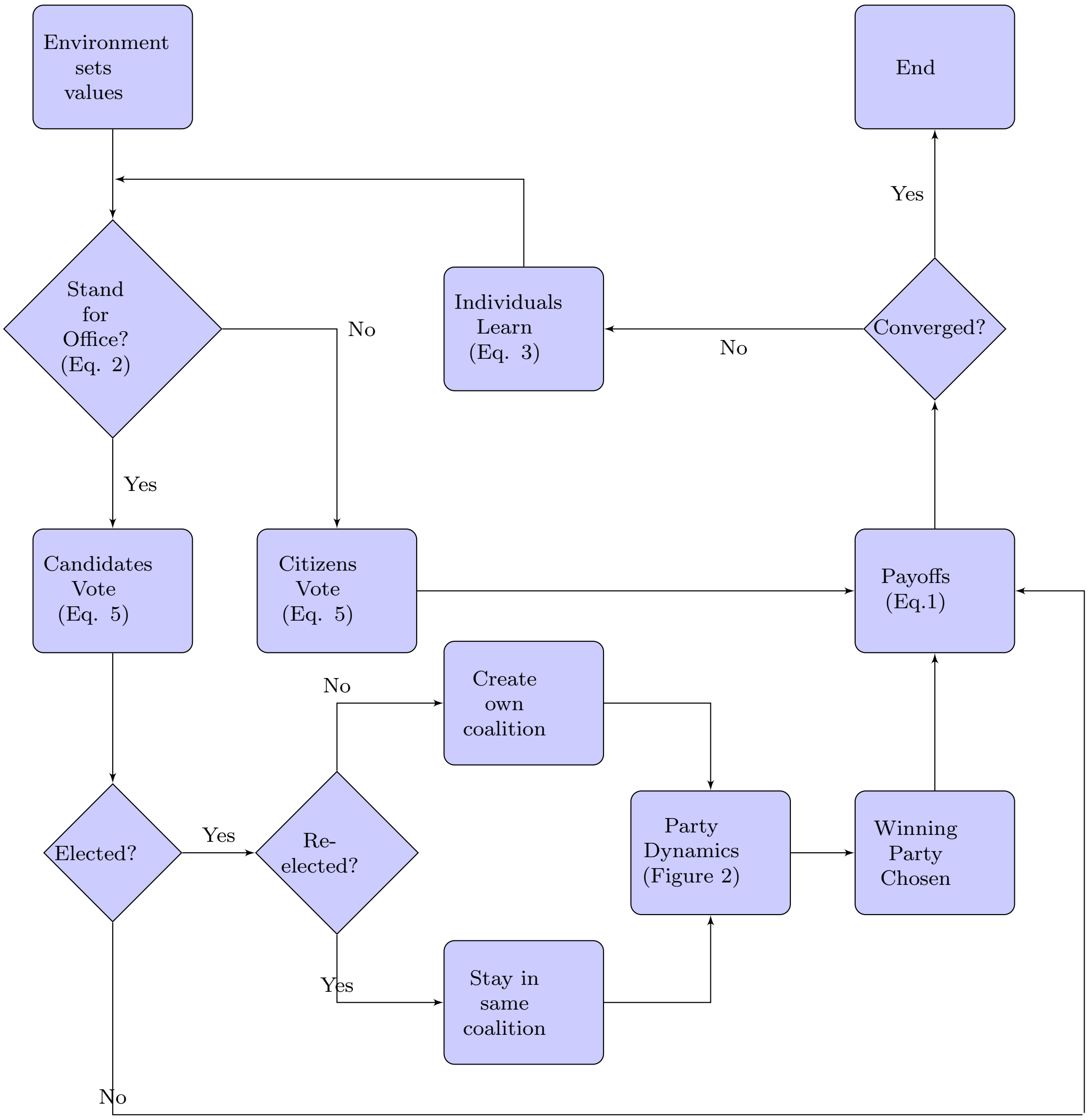


Figure 1: Flowchart depicting the order of a citizen's choices within the model.

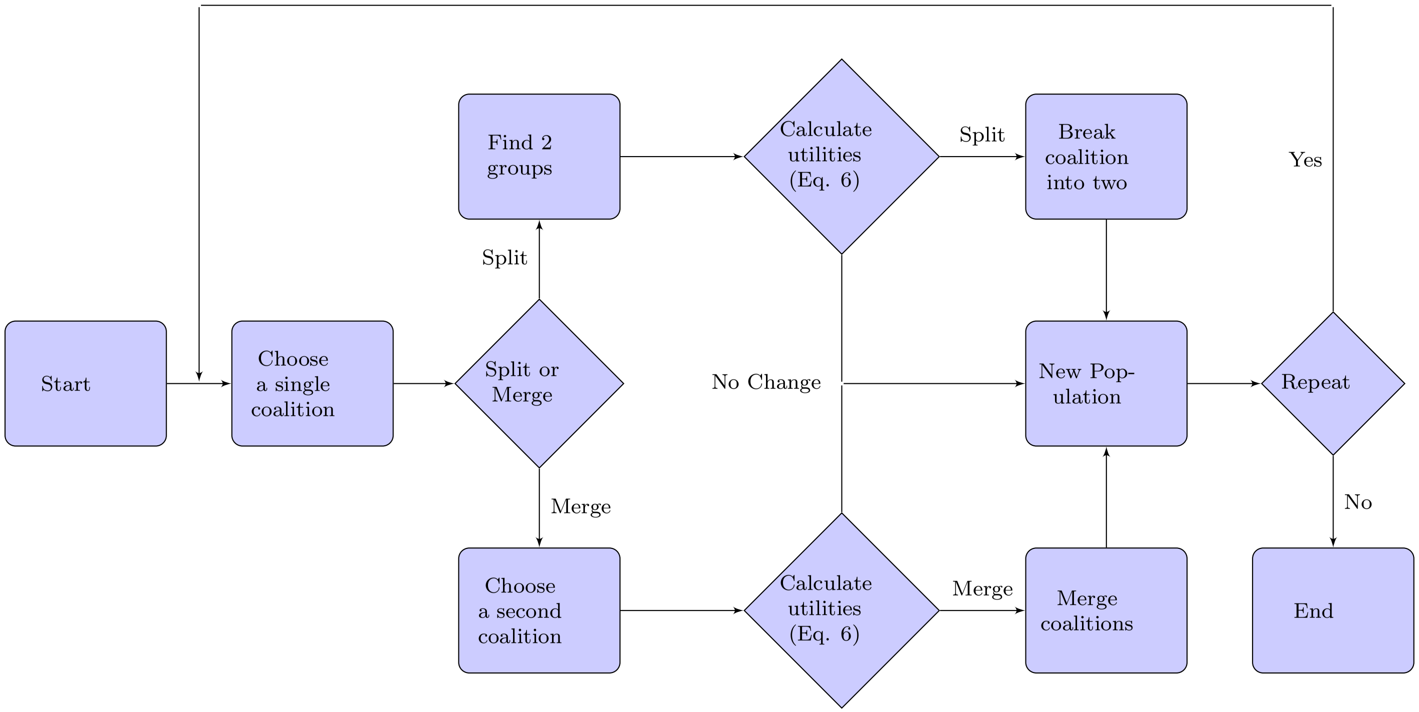


Figure 2: Flowchart detailing the order of events in the process of splitting and merging coalitions.

B Details of algorithms

B.1 The Gram-Schmidt scheme

As discussed in the paper, predicting the relative positioning of political parties has been the subject of considerable attention at least since Downs (1957). One difficulty in approaching this question in the context of multiple policy dimensions is how to display, conceptualize, and compare results. Many landscapes are reflections or rotations of others. Given that we attach no substantive interpretation to the policy dimensions, these differences are not of much interest. Our approach is to consider the positions of each party relative to the largest party. To do this a Gram-Schmidt scheme (as described by Golub and Van Loan (1996)) is employed to produce an orthonormalization of the set of vectors describing party positions. We first define the location of the largest party to be the origin of a new coordinate system. From here a series of M orthogonal vectors, v^i for $i = 1 - M$ are calculated, corresponding to the axis of the new coordinate space such that for the Q^{th} largest party which has position p^Q, $v^i \dot{p}^Q = 0$ for all $i \geq Q$ and where $M \leq N$ where N was the dimensionality of the original coordinate space.

We now consider an example of how the process works: The results of three simulations carried out in two dimensions are shown in figure B.1. In each case there are three parties distributed within the policy space, however, beyond this it is not possible directly to identify any similarities between the configurations. Figures B.1 shows the party locations after Gram-Schmidt transformation. The largest party in each case is now located at the origin with the second largest on the X-axis, all parties within each simulation maintain their relative positions, however, comparison across simulation is simplified. It can now be seen that two of the result sets are similar in the patterns of parties whilst the third differs significantly. The k-means algorithm for 2 clusters successfully identifies these (figure B.1). The first being based on the triangle and cross results whilst the second represent solely the circle results.

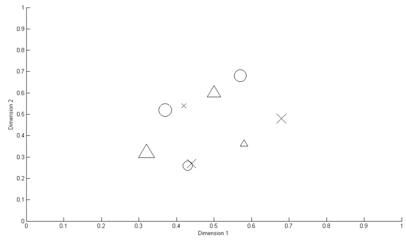


Figure 3: Results of three simulations, each marker is a party at the end of the simulation with markers of the same type coming from the same simulation. Marker size corresponds to party size

.

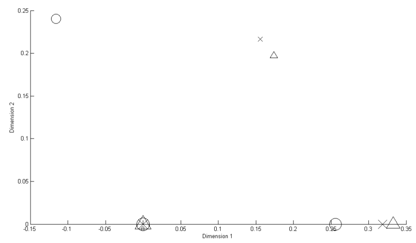


Figure 4: Results of three simulations shown in figure 1 after the application of the Gram-Schmidt scheme.

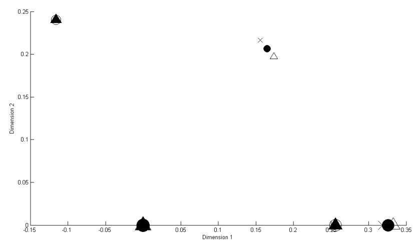


Figure 5: Results of three simulations shown in figure 1 after the application of the Gram-Schmidt scheme with two equilibria represented by filled shapes found by the k means algorithm.

B.2 Equilibrium Identification

To do this a Gram-Schmidt scheme is employed to produce an orthonormalization of the set of vectors describing party positions. It is worth noting that one consequence of the Gram-Schmidt scheme is that equilibria that are reflections or rotations of another before normalization are equivalent afterwards. We perform this orthnormalisation for the results of all simulations and analyse the equilibria.

There are two different sources for variation in equilibria. Firstly, different equilibria arise due to differences in the preferences of citizen-candidates as determined here by the random number seed. These differences are expected in any such model, computational or otherwise. The second source of variation is due to path dependence. For example, if there were two citizen-candidates in a particular district with extremely similar preferences, it may be that it is in the interests of both for one but not both of them to stand. In our model, provided that they receive similar amounts of utility from standing, which of the two stands in equilibrium is potentially path-dependent. That is, there are two possible stable outcomes (in this model) one candidate stands with probability 1 and one candidates stands with probability 0. But, which of the two is which may be dependent, for example, on which is the first to randomly stand when the other doesn't. The minor differences in equilibria for reasons such as this are not as interesting as larger qualitative differences between equilibria, which result in different sized or located parties.

The distinction made above is an imperfect one, whether two equilibria count as being qualitatively different is in part subjective. As such, similarly to section A.3, we employ a statistical approach to find the number of distinct clusters in the data. The data for each parameter combinations, are the results of 1000 repetitions of the simulation with different random seeds. As before, the results from each simulation are converted to a set of vectors in which each vector contains the policy of a party. A Gram-Schmidt process is applied to these vectors and the results combined with the relative party sizes to produce a single vector for each simulation characterizing its results. A k-means clustering algorithm is applied to the set of 1000 vectors for a range of values of k, to identify clusters of almost identical equilibria. For each value, 1000 repetitions of the algorithm are run and the minimum value of k required to explain 90% of the total variance is found.[7] Accordingly, we define the number of equilibria as, k, the number of distinct clusters identified.[8]

B.3 The form and number of equilibria

Table 2 reports the number of equilibria identified for each combination of parameters. One immediate conclusion is that the number of equilibria is relatively small. Most important is that the number of equilibria is always considerably smaller than the number of observations for any given cell. This suggests we aren't just sampling points on a continuum of possible equilibria. A possible concern is that this is an artifact of the algorithm we employ, but in fact the results are remarkably insensitive to the details of the procedure. The most notable result is that there is step change in the number of equilibria associated with the move from a single dimension (henceforth *dim*) to higher dimensional spaces. The electoral system, henceforth *pr*, seems to have no impact. These results are slightly surprising – one might have expected there to be more possible equilibria in higher dimensional spaces but instead the converse is true. Results (not reported) for the number of equilibria in individual constituency simulations, as described in Appendix A, show that we find many more equilibria in the absence of parties. One interpretation of this is that the additional structure imposed by considering multiple constituencies, linked by the formation of parties, reduces the set of feasible equilibria considerably.

One issue is whether we identify all or just some of the possible equilibria. The model is more convincing if we may be confident that the results describe all, or at the very least the vast majority, of the potential outcomes. To alleviate such concerns we simulated the case of $dim = 2$, $pr = 2$ for, $50,000$, iterations.

[7] The results are not sensitive to the choice of the percentage of variance explained.

[8] An augmented algorithm in which small perturbations of party sizes were applied to identify seemingly different but in effect identical equilibria was applied, but the results don't change meaningfully.

Table 1: Number of Equilibria Identified by Number of Simulations

Number of Simulations	100	200	500	1000	2000	5000	10000	20000	50000
Number of Equilibria	8	8	9	9	9	9	9	9	9

The results suggest that the algorithm, even for relatively few simulations identifies almost all of the equilibria. A 100-fold increase in the number of random number seeds, and thus distributions of individual preferences etc., identify no more equilibria.[9] As such, we argue that the results below can be seen as a comprehensive representation of the properties of the model.

Table 2: Number of Equilibria Identified

Dimension	1	2	3	4
PR	Number of Equilibria			
1	86	5	6	9
2	99	9	7	4
3	50	48	17	7
4	22	15	9	12
5	37	27	14	17
6	41	28	13	19
8	47	32	15	18
10	46	32	17	21
12	52	30	18	20
15	48	30	15	15
20	48	27	14	8
24	48	27	18	6
30	47	31	19	5
40	48	26	26	4
60	47	21	22	4
120	47	28	20	7

B.4 The effect of dimension on party effectiveness

Consider two coalitions of size n_1 and n_2 located in a policy space. In one dimension we take the centres of each party to be the points 0 and 1. We assume that half of each coalitions members are located to the left of the centre and half to the right. In two dimensions we take the centers of each party to be the points $(0, 0)$ and $(1, 0)$ located in the space bounded by $(0, -1), (0, 1), (1, -1), (1, 1)$. As before we assume that half of

[9]This is not a claim that no more equilibria are found, but that these are not distinct up-to an orthonormalization, from those already identified.

each coalition's members are located to the left of the center and half to the right but all lie within the y-range.[10] In both cases the total number of members within the space is $\frac{n_1}{2} + \frac{n_2}{2}$.

We define the area of a party to be the volume of space in which an individual would prefer to be in that party over the other. Figure B.4 shows for both 1 and 2 dimensions the change in size of party 1's area when one individual moves from party 2 to party 1. The figure also shows the average area occupied by each individual in the space (I.E. $\frac{1}{\#members}$).[11] In one dimension the change in area is always less than the average area occupied by an individual. This means that when an individual changes party, on average both parties will be stable. In contrast in two dimensions, for a large range of party sizes, one individual changing party will change the area of the party by more than the average area occupied by an individual. As a result this is likely to lead to another doing the same. A party may, therefore, slowly draw members from another whereas in one dimension this is not the case.

There are obviously simplifications in this model - for example the constrained space and uniform distribution of candidates - however, it illustrates the basic mechanism at work.

References

Degan, Arianna and Antonio Merlo, "Do voters vote ideologically?," *Journal of Economic Theory*, 2009, *144* (5), 1868 – 1894.

Downs, Anthony, *An Economic Theory of Democracy*, New York: Harper & Row, 1957.

Eguia, Jon X, "Challenges to the Standard Euclidean Spatial Model," in "Advances in Political Economy," Springer Berlin Heidelberg, 2013, pp. 169–180.

Eguia, Jon X., "On the spatial representation of preference profiles," *Economic Theory*, 2013, *52* (1), 103–128.

Golub, Gene H. and Charles F. Van Loan, *Matrix Computations*, 3rd ed., John Hopkins University Press, 1996.

Hartigan, J. A. and M. A. Wong, "A K-means clustering algorithm," *Applied Statistics*, 1979, *28*, 100–108.

Humphreys, Macartan and Michael Laver, "Spatial models, cognitive metrics, and majority rule equilibria," *British Journal of Political Science*, 2010, *40* (01), 11–30.

[10]The result holds for a wide range of y dimension sizes.

[11]We implicitly assume that individuals are uniformly distributed, however, if a greater proportion of individuals lie within the space the results do not qualitatively change.

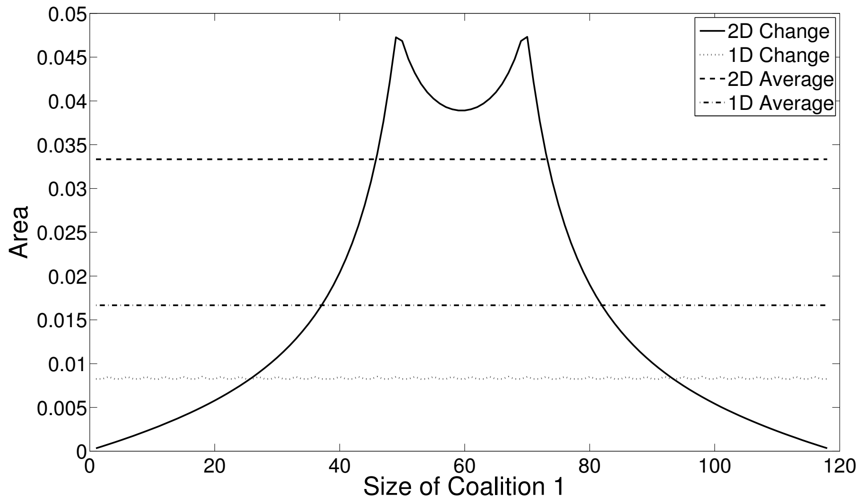


Figure 6: The effect on other party members of one individual moving for one and two dimensional spaces

Lloyd, Stuart P., "Least Squares Quantitization in PCM," *IEEE Transactions on Information Theory*, 1982, *28* (2), 129–137.

Okabe, Atsuyuki, Barry Boots, Kokichi Sugihara, and Sung Nok Chiu, *Spatial tessellations: concepts and applications of Voronoi diagrams*, Vol. 501, Wiley. com, 2009.

Osborne, Martin J. and Al Slivinski, "A Model of Political Competition with Citizen-Candidates," *Quarterly Journal of Economics*, Febuary 1996, *111* (1), 65–96.

Ray, Debraj and Rajiv Vohra, "Equilibrium binding agreements," *Journal of Economic Theory*, 1997, *73* (1), 30–78.

Rustichini, Aldo, "Optimal Properties of Stimulus-Response Learning Models.," *Games and Economic Behavior*, October 1999, *29* (1-2), 244–273.