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:

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

   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.

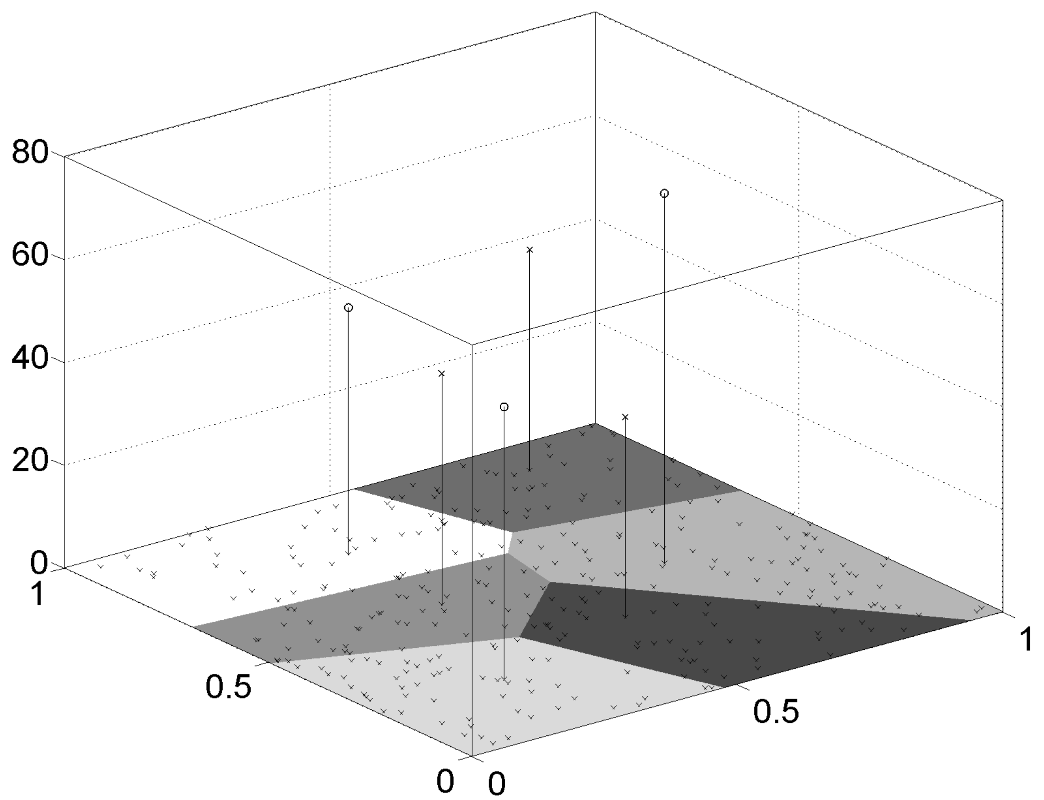
4. The process is repeated for a large number of iterations until an equilibrium has been reached. This is defined as where $P_{jt}^{stand} = \{0, 1\}\forall j$, that is every citizen either stands or not with certainty. This rules out outcomes analogous to mixed strategy equilibria in which some citizens stand with probability $0 < P_{jt}^{stand} < 1$ 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:

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

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

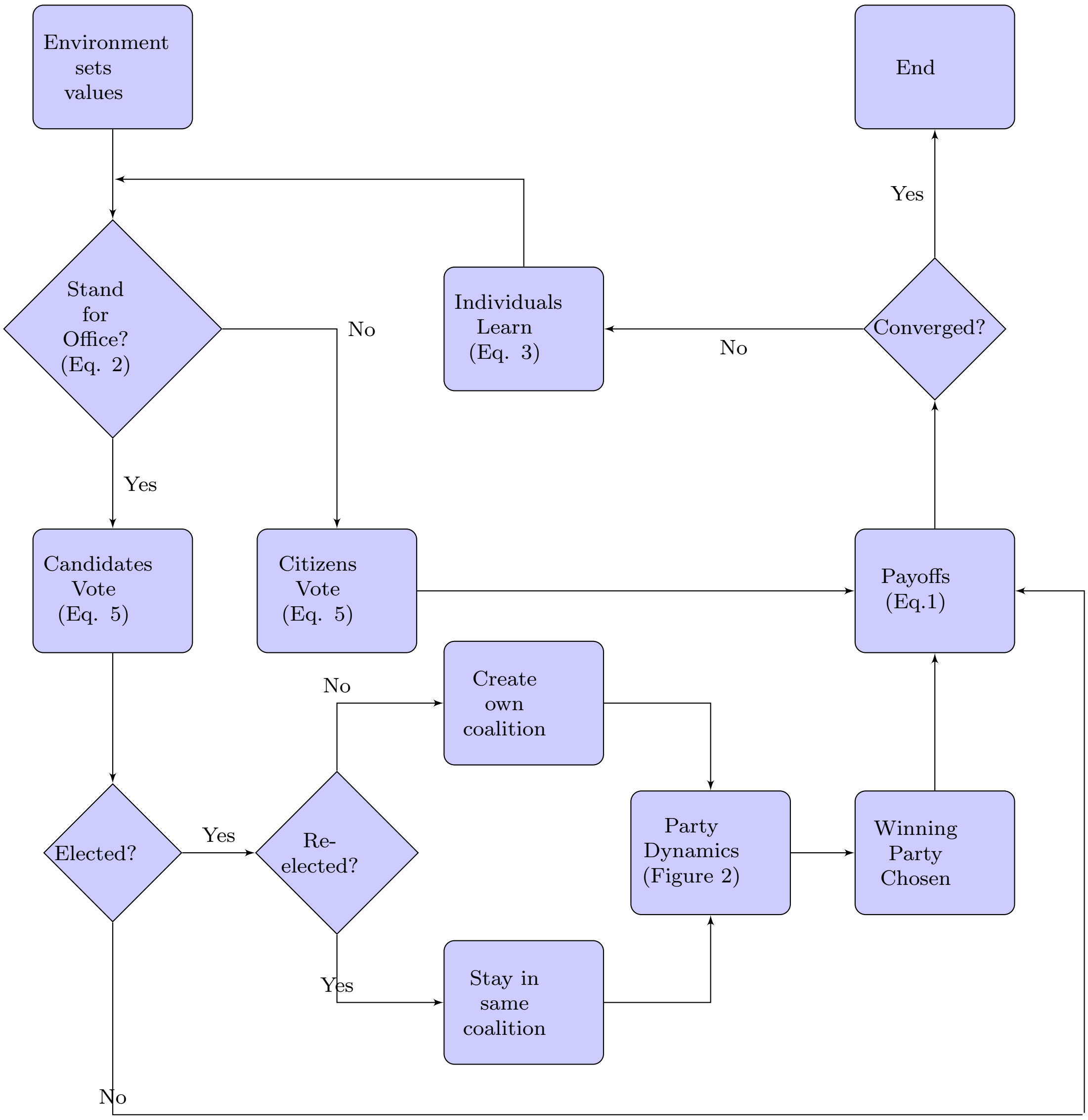


Figure 1: Flowchart depicting the order of a citizen's choices within the model.

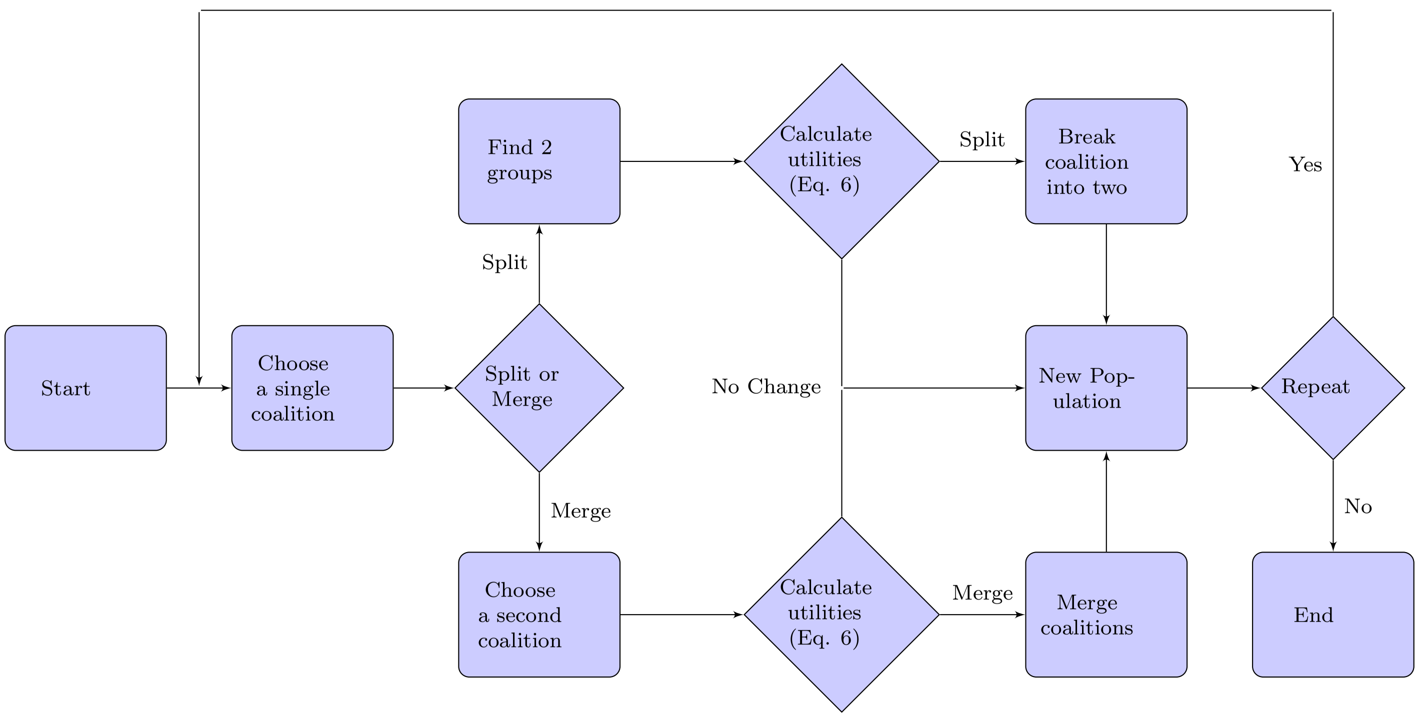


Figure 2: Flowchart detailing the order of events in the process of splitting and merging coalitions.

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.

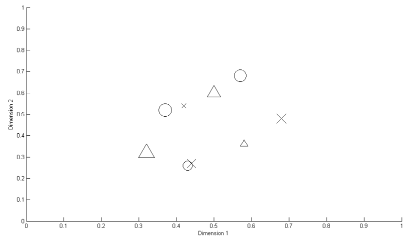


Figure 3: Results of three simulations, each marker is a party at the end of the simulation with markers of the same type coming from the same simulation. Marker size corresponds to party size

.

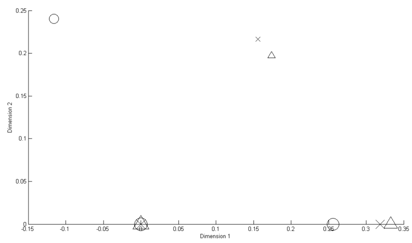


Figure 4: Results of three simulations shown in figure 1 after the application of the Gram-Schmidt scheme.

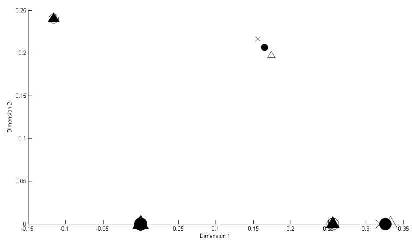


Figure 5: Results of three simulations shown in figure 1 after the application of the Gram-Schmidt scheme with two equilibria represented by filled shapes found by the k means algorithm.

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.

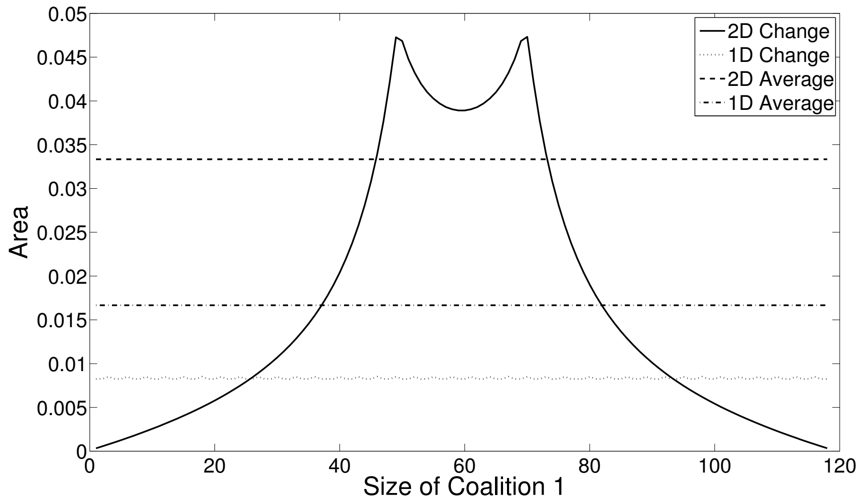


Figure 6: The effect on other party members of one individual moving for one and two dimensional spaces

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.