14. Divergence in the Spatial Stochastic Model of Voting

Norman Schofield
Center in Political Economy, Washington University in St. Louis, USA

[I]t may be concluded that a pure democracy, by which I mean a society, consisting of a small number of citizens, who assemble and administer the government in person, can admit of no cure for the mischiefs of faction … Hence it is that democracies have been spectacles of turbulence and contention; have ever been found incompatible with personal security … and have in general been as short in their lives as they have been violent in their deaths.

A republic, by which I mean a government in which the scheme of representation takes place, opens a different prospect...

[I]f the proportion of fit characters be not less in the large than in the small republic, the former will present a greater option, and consequently a greater probability of a fit choice.

— *James Madison*, 1787 (quoted in Rakove 1999).

1. Introduction

Much research has been devoted over the last few decades in an attempt at constructing formal models of political choice in electoral systems based on proportional representation (PR). In large degree these models have been most successful in studying the post-election phase of coalition bargaining (Laver and Schofield 1990; Laver and Shepsle 1996; Banks and Duggan 2000). Such models can take the locations and strengths of the parties as given. Attempts at modelling the electoral phase have met with limited success, since they have usually assumed that the policy space is restricted to one dimension, or that there are at most two parties. The extensive formal literature on two party electoral competition was typically based on the assumption that parties or candidates adopted positions in order to win (Calvert 1985.) In PR systems it is unlikely that a single party can gain enough votes to win outright (Cox 1990, 1997; Strom 1990).

However, there is a well developed empirical modelling technique which studies the relationship between party positions and electoral response.

Dating back to empirical models of U.S. elections (Poole and Rosenthal 1984) and continuing more recently with Quinn, Martin and Whitford (1999), Alvarez, Nagler and Bowler (2000), Alvarez, Nagler and Willette (2000), such models elucidate the relationship between voter preferred positions, party positions, sociodemographic variables and voter choice. Given such an empirical model, it is possible to perform a 'counter-factual experiment' to determine the voter response if economic conditions or party positions had been different (Glasgow and Alvarez 2003). It is then natural to seek the existence of 'Nash equilibria' in the empirical model – a set of party positions from which no party may deviate to gain advantage. Since the 'utility functions' of parties are, in fact unknown, I can use the 'counter factual experiment' to make inferences about the political game. That is, knowing the relationship between positions and electoral outcome, I make the hypothesis that the actual positions are indeed Nash equilibria in some formal political game. In principle, this could give information about the unknown utility functions of political leaders.

The most obvious assumption to make about party utility functions is that they can be identified with the vote shares of the parties. In PR systems, in particular, vote shares and seat shares are approximately identical. At least for some parties, increasing seat shares will increase the probability of membership of a governing coalition, thus giving access to government perquisites and the opportunity to affect policy. Since empirical models necessarily have a stochastic component, the natural formal tool to use in seeking Nash equilibria is the so-called probabilistic spatial model of voting. Such models have an inherent stochastic component associated with voter choice. Developing the earlier argument of Hinich (1977), Lin, Enelow and Dorussen (1999) have asserted the 'mean voter theorem' that all vote maximizing parties should converge to the electoral mean (Hinich 1982, 1984*a,b*, 1989*a,b*; Coughlin 1994; Adams 1999*a,b*; 2001; Adams and Merrill 2000, 2001; Banks and Duggan 2005; and McKelvey and Patty 2006).

Since the early exposition of Duverger (1954), ample evidence has been collected that indicates that parties do not converge in this way (Daalder 1984; Budge, Robertson, and Hearl 1987; Rabinowitz, MacDonald and Listhaug 1991). The contradiction between the conclusions of the formal model and observation has led to efforts to modify the voter model by adding 'party identification' or 'directional voting' (Merrill, Grofman, and Feld 1999) However, such attempts have not made it obvious why small parties apparently adopt 'radical' positions (Adams and Merrill 1999, 2000).

An alternative way to modify the basic formal model is to add the effect of 'valence.' Stokes (1963, 1992) first introduced this concept many years ago. 'Valence' relates to voters' judgements about positively or negatively evaluated conditions which they associate with particular parties or candidates. These judgements could refer to party leaders' competence, integrity, moral stance or 'charisma' over issues such as the ability to deal with the economy, foreign threat, etc. The important point to note is that these individual judgements are independent of the positions of the voter and party. Estimates of these judgements can be obtained from survey data – see, for example, the work on Britain by Clarke, Stewart and Whiteley (1997, 1998) and Clarke et al. (2004). However, from such surveys it is difficult to determine the 'weight' that an individual voter attaches to the judgement in comparison to the policy difference. As a consequence, the empirical models usually estimate valence for a party or party leader as a constant or intercept term in the voter utility function. The party valence variable can then be assumed to be distributed throughout the electorate in some appropriate fashion. This stochastic variation is expressed in terms of 'disturbances,' which, in the most general models, can be assumed to be distributed multivariate normal with a non-diagonal covariance matrix. This formal assumption parallels that of multinomial probit estimation (MNP) in empirical models (Quinn, Martin, and Whitford 1999). A more restrictive assumption is that the errors are independently and identically distributed by the Type I extreme value, or log Weibull, distribution (Quinn and Martin 2002; Dow and Endersby 2004). This parallels multinomial logit estimation (MNL) in empirical estimation.

This conception of voting as judgement as well as the expression of voter interest or preference is clearly somewhat similar to Madison's understanding of the nature of the choice of Chief Magistrate in the Republic. Madison's argument may well have been influenced by Condorcet's work on the so-called 'Jury Theorem' (Condorcet 1785; Schofield 2006). However, Madison's conclusion about the 'probability of a fit choice' depended on assumptions about the independence of judgements from interests, and these assumptions are unlikely to be valid. Condorcet's work has recently received renewed attention (McLennan 1998) and formal models have been presented based on the notion of valence (Ansolabehere and Snyder 2000; Groseclose 2001; Aragones and Palfrey 2002, 2005). However little work has been done on developing the stochastic voter model when judgements, or valences, and preferences are both involved. (An important recent exception is Adams and Merrill (2005)).

In a sense, the above work and the results presented here can be seen as a contribution to the development of Madison's conception of elections in representative democracies as methods of aggregration of both preferences and judgements. Unlike Madison, however, these models also concern themselves with the response of elected representatives to electoral perceptions. Indeed, the most important contribution of such models is the determination of conditions under which heterogeneity can be maintained. Although the formal model is applied to electoral systems based on proportional representation, it could in principle be applied to electoral systems based on plurality, such as in Britain or the U.S.

2. Modelling Elections

Formal models of voting usually make the assumption that political agents, whether parties or candidates, attempt to maximize expected vote shares (or, in two party contests, the plurality). The 'stochastic' version of the model typically derives the 'mean voter theorem' that each agent will adopt a 'convergent' policy strategy at the mean of the electoral distribution. This conclusion is subject to a constraint that the stochastic component is 'sufficiently' important. (Lin, Enelow, and Dorussen 1999). Because of the apparent inconsistency between the theory and empirical evidence, I shall re-examine the implications of the formal stochastic model when there are significant valence differences between the candidates or party leaders.

The 'valence' of each party derives from the average weight, given by members of the electorate, to the overall competence of the particular party leader. In empirical models, a party's valence is independent of its current policy declaration, and can be shown to be statistically significant in the estimation. Because valence may comprise all those non-policy aspects of the competition, I attempt to account for as many such factors as possible by including sociodemographic features in the model.

As the analysis shows, when valence terms are incorporated in the model, then the convergent vote maximizing equilibrium may fail to exist. I contend that the empirical evidence is consistent with a formal stochastic model of voting in which valence terms are included. Low valence parties, in equilibrium, will tend to adopt positions at the electoral periphery. High valence parties will contest the electoral center, but will not, in fact, occupy the electoral mean. I use evidence from elections in Israel (Schofield et al. 1998; Schofield and Sened 2005*a*) to support and illustrate this argument.

For the discussion and analysis of the case of Israel I combine available and original survey data for Israel for 1988 to 1996, that allows me to construct an empirical model of voter choice in Knesset elections. I use expert evaluations to estimate party positions and then construct an empirical vote model that I show is statistically significant. Using the parameter estimates of this model, I develop a 'hill climbing' algorithm to determine the empirical equilibria of the vote-maximizing political game. Contrary to the conclusions of the formal stochastic vote model, the 'mean voter' equilibrium, where all parties adopt the same position at the electoral mean, did not appear as one of the simulated equilibria. Since the voter model that I develop predicts voter choice in a statistically significant fashion, I infer that the assumptions of the formal stochastic vote model are compatible with actual voter choice. Moreover, equilibria determined by the simulation were 'close' to the estimated configuration of party positions for the three elections of 1988, 1992 and 1996. I infer from this that the assumption of vote share maximization on the part of parties is a realistic assumption to make about party motivation.

To evaluate the validity of the 'mean voter theorem' I consider, a formal vote model. The usual assumption to make to ensure existence of a 'Nash equilibrium' at the mean voter position depends on showing that all party vote share functions are 'concave' in some domain of the party strategy spaces (Banks and Duggan 2005). Concavity of these functions depends on the parameters of the model. Because the appropriate empirical model for Israel incorporated valence parameters, these were part of the concavity condition for the baseline formal model. Concavity is a global property of the vote share functions, and is generally difficult to empirically test. I focus on a weaker equilibrium property, that of 'local Nash equilibrium,' or LNE. Necessary and sufficient conditions for existence of an LNE at the electoral mean can be obtained from 'local concavity' conditions on the second derivative (the Hessian) of the vote share functions. If local concavity fails, then so must concavity. The necessary condition required for existence of LNE at the origin in the formal vote model is violated by the estimated values of the parameters in the empirical model in these elections in Israel. Consequently, the empirical model of vote maximizing parties could not lead us to expect convergent strategies at the mean electoral position. The electoral theorem presented below is valid in a policy space of unrestricted dimension, but has a particularly simple expression in the two-dimensional case.

The theorem allows me to determine whether the mean voter position is a best response for a low valence party when all other parties are at the mean. In the empirical model I estimate that low valence parties would, in fact, minimize their vote share if they chose the mean electoral position. This inference leads me to the following conclusions (i) some of the low valence parties, in maximizing vote shares, should adopt positions at the periphery of the electoral distribution (ii) if this does occur, then the first order conditions for equilibrium, associated with high valence parties at the electoral mean, will be violated. Consequently, for the sequence of elections in Israel, we should expect that it is a non-generic property for any party to occupy the electoral mean in any vote maximizing equilibrium.

Clearly, optimal party location depends on the valence by which the electorate, on average, judges party competence. The simulations suggest that if a single party has a significantly high valence, for whatever reason, then it has the opportunity to locate itself near the electoral center. On the other hand, if two parties have high, but comparable valence, then the simulation suggests that neither will closely contest the center. We can observe that the estimated positions of the two high valence parties, Labor and Likud, are almost precisely identical to the simulated positions under expected vote maximization. The positions of the low valence parties are, as predicted, close to the periphery of the electoral distribution. However they are not identical to simulated vote maximizing positions. This suggests that the perturbation away for vote maximizing equilibria is either due to policy preferences on the part of party principals or to the effect of party activists (Aldrich 1983*a*,*b*; Miller and Schofield 2003).

The formal and empirical analyses presented here are applicable to any polity using an electoral system based on proportional representation. The underlying formal model is compatible with a wide variety of different theoretical political equilibria. The theory is also compatible with the considerable variation of party political configurations in multiparty systems (Laver and Schofield 1998).

The analysis of the formal vote model emphasizes the notion of 'local' Nash equilibrium in contrast to the notion of a 'global' Nash equilibrium usually employed in the technical literature. One reason for this emphasis is that I deploy the tools of calculus and simulation via hill climbing algorithms to locate equilibria. As in calculus, the set of local equilibria must include the set of global Nash equilibria. Sufficient conditions for existence of a global Nash equilibrium are therefore more stringent than for a local equilibrium. In fact, the necessary and sufficient condition for a local equilibrium at the electoral center, in the vote maximizing game with valence, is so stringent that I regard it to be unlikely to obtain in polities with numerous parties and varied valences. I therefore infer that existence of a global Nash equilibrium at the electoral center is very unlikely in such polities. In contrast, the sufficient condition for a local, non-centrist equilibrium is much less stringent. Indeed, in each polity there may well be multiple local equilibria. This suggests that the particular configuration of party positions in any polity can be a matter of historical contingency.

3. Empirical Estimation of Elections in Israel 1988–1996

I assume that the political preferences (or beliefs) of voter i can be described by a 'latent' utility vector of the form

$$\mathbf{u}_i(x_i,\mathbf{z}) = (u_{i1}(x_i,z_1),\ldots,u_{ip}(x_i,z_p)) \in R^p. \tag{1}$$

Here $\mathbf{z}=(z_1,\ldots,z_p)$ is the vector of strategies of the set, P, of political agents (candidates, parties, etc.). The point z_j is a vector in a policy space X that I use to characterize party j. (For the formal theory it is convenient to assume X is a compact convex subset of Euclidean space of dimension w, but this is not an absolutely necessary assumption. I make no prior assumption that $w=1$.) Each voter, i, is also described by a vector x_i, in the same space X, where x_i, is used to denote the beliefs or 'ideal point' of the voter. I assume

$$u_{ij}(x_i,z_j) = \lambda_j - A_{ij}(x_i,z_j) + \theta_j^{\mathrm{T}}\eta_i + \varepsilon_j. \tag{2}$$

I use $A_{ij}(x_i,z_j)$ to denote some measure of the distance between the vectors x_i and z_j. I follow the assumption of the usual 'Euclidean' model and assume that $A_{ij}(x_i, z_j) = \beta \| x_i - z_j \|^2$ where $\| \cdot \|$ is the Euclidean norm on X and β is a positive constant. It is also possible to use an ellipsoidal distance function for A_{ij}, although in this case the necessary and sufficient condition for equilibrium has a more complex expression. The term λ_j is called *valence* and has been discussed above. The *k*-vector θ_j represents the effect of the k different sociodemographic parameters (class, domicile, education, income, etc.) on voting for the party *j* while η_i is a *k*-vector denoting the i^{th} individual's relevant 'sociodemographic' characteristics. I use θ_j^{T} to denote the transpose of θ_j so $\theta_j^{\mathrm{T}}\eta_i$ is a scalar. The abbreviation SD is used throughout to refer to models involving sociodemographic characteristics The vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_j, \ldots, \varepsilon_p)$ is the 'stochastic' error , where ε_j is associated with the j-th party. Some recent analyses have assumed that ε is multivariate normal, and used Markov Chain Monte Carlo (MCMC) methods, allowing for multinomial probit (MNP) estimation when the errors are covariant (Chib and Greenberg 1996; Schofield et al. (1998); Quinn, Martin, and Whitford 1999). Assuming that the errors are independent and identically distributed via the Type I extreme value (or log-Weibull) distribution gives a multinomial logit (MNL) model. The details of this modelling assumption are given in the next section, where I present formal results based on this assumption. MNP models are generally preferable because they do not require the restrictive assumption of 'independence of irrelevant alternatives' (Alvarez and Nagler 1998). I use a MNL model in this paper because comparison of MNL and MNP models suggest that the simpler MNL model gives an adequate account of voter choice (Dow and Endersby (2004: 111). It is also much easier to use the MNL empirical model to simulate vote-maximizing strategies by parties (Quinn and Martin 2002).

I shall now show that divergence results simply from the hypothesis of vote maximizing in the context of the empirical stochastic model, based on Type I extreme value distribution when valence is important.

Table 1 gives the results for the five elections in Israel from 1988 to 2003, while Table 2 gives details of the factor model for 1996. Table 3 gives the estimation details for the election of 1996, showing the spatial, valence and sociodemographic coefficients. Similar estimations were obtained for the elections of 1992 and 1988. These estimations are based on factor analysis of the surveys conducted by Arian and Shamir (1990, 1999, 1995) for the three elections. Party positions for the three election years 1988, 1992 and 1996 were estimated by expert analysis of party manifestos, using the same survey questionnaires and these are shown in figures 1, 2 and 3. These figures also show the 'smoothed' distributions of voter ideal points for 1996 and 1992 and 1988. (The outer contour line in each figure contains 95% of the voter ideal points).

Table 1. Israeli elections 1998–2003

	Party	Knesset Seats				
		1988	1992	1996	1999	2003
Left	Labor (LAB)	39	44	34	28	19[a]
	Democrat (ADL)	1	2	4	5	2[a]
	Meretz,Ratz (MZ)	–	12	9	10	6
	CRM, MPM,PLP	9	–	–	–	3
	Communist (HS)	4	3	5	3	3
	Balad	–	–	–	2	3
Subtotal		53	61	52	48	36
Center	Olim	–	–	7	6	2[b]
	III Way	–	–	4	–	–
	Center	–	–	–	6	–
	Shinui (S)	2	–	–	6	15
Subtotal		2		11	18	17[b]
Right	Likud (LIK)	40	32	30	19	38[b]
	Gesher	–	–	2	–	–
	Tzomet (TZ)	2	8	–	–	–
	Yisrael beiteinu	–	–	–	4	7
Subtotal		42	40	32	23	45
Religious	Shas (SHAS)	6	6	10	17	11
	Yahadut (AI, DH)	7	4	4	5	5
	(Mafdal) NRP	5	6	9	5	6
	Moledet (MO)	2	3	2	4	–
	Techiya,Thia (TY)	3	–	–	–	–
Subtotal		23	19	25	31	22
Total		120	120	120	120	120

[a] Am Ehad or ADL under Peretz, combined with Labor, to give the party $19+2=21$ seats.
[b] Olim joined Likud to form one party giving Likud $38+2=40$ seats, and the right $40+7=47$ seats.

Table 2. Factor analysis for the 1996 Israeli election

Issue question	Factor weights	
	Security	Religion
Chance for peace	0.494 (0.024)	–
Land for peace	0.867 (0.013)	–
Religious law vs. Democracy	0.287 (0.038)	0.613 (0.054)
Must stop peace process	–0.656 (0.020)	–
Agreement with Oslo accord	–0.843	–

Table 2. Factor Analysis, 1996 Israeli 1996 Election (cont'd)

Issue question	Factor weights	
	Security	Religion
Oslo accord contributed to safety	–0.798 (0.016)	–
Personal safety after Oslo	–0.761 (0.020)	–
Israel should talk with PLO	–0.853 (0.016)	–
Opinion of settlers	0.885 (0.015)	–
Agree that Palestinians want peace	–0.745 (0.016)	–
Peace agreement will end Arab-Israeli conflict	–0.748 (0.018)	–
Agreement with Palestinian state	–0.789 (0.016)	–
Should encourage Arabs to emigrate	0.618 (0.022)	–
Israel can prevent war	–0.843 (0.019)	–
Settlements 1	0.712 (0.014)	–
Settlements 2	0.856 (0.014)	–
National security	0.552 (0.023)	–
Equal rights for Arabs and Jews in Israel	–0.766 (0.018)	–
More government spending towards religious institutions	0.890 (0.035)	–
More government spending towards security	0.528 (0.049)	0.214 (0.065)
More government spending on immigrant absorption	0.342 (0.052)	0.470 (0.065)
More government spending on settlements	0.597 (0.040)	0.234 (0.055)
More government spending in Arab sector	–0.680 (0.019)	–
Public life should be in accordance with Jewish tradition		1.000 constant
Views toward an Arab minister	–0.747 (0.019)	
Var (security)	1.000	constant
Var (religion)	0.732	
Covar (security, religion)	0.591	

Table 3. Multinomial Logit Analysis,1996 Israeli Election

	Party	Posterior Mean	95% Confidence Interval Lower Bound	Upper Bound
Spatial Distance	β	1.117	0.974	1.278
Constant	Shas	–2.960	–7.736	1.018
	Likud	3.140	0.709	5.800
	Labor	4.153	1.972	6.640
	NRP	–4.519	–8.132	–1.062
	Moledet	–0.893	–4.284	2.706
	III Way	–2.340	–4.998	0.411
Ashkenazi	Shas	0.066	–1.831	1.678
	Likud	–0.625	–1.510	0.271
	Labor	–0.219	–0.938	0.492
	NRP	1.055	–0.206	2.242
	Moledet	0.819	–0.560	2.185
	III Way	–0.283	–1.594	1.134
Age	Shas	0.014	–0.058	0.086
	Likud	–0.024	–0.063	0.008
	Labor	–0.040	–0.077	–0.012
	NRP	–0.064	–0.111	–0.020
	Moledet	–0.026	–0.088	0.026
	III Way	0.014	–0.034	0.063
Education	Shas	–0.377	–0.693	–0.063
	Likud	–0.032	–0.180	0.115
	Labor	0.011	–0.099	0.120
	NRP	0.386	0.180	0.599
	Moledet	0.049	–0.219	0.305
	III Way	–0.067	–0.298	0.150
Religious Observation	Shas	3.022	1.737	4.308
	Likud	0.930	0.270	1.629
	Labor	0.645	0.077	1.272
	NRP	2.161	1.299	3.103
	Moledet	0.897	–0.051	1.827
	III Way	0.954	0.031	1.869
Correctly Predicted	Shas	0.309	0.210	0.414
	Likud	0.707	0.672	0.740
	Labor	0.717	0.681	0.752
	NRP	0.408	0.324	0.493
	Moledet	0.078	0.046	0.115
	III Way	0.029	0.017	0.043
	Meretz	0.286	0.226	0.349
	Entire Model	0.638	0.623	0.654
$n = 791$		Log marginal likelihood: –465.		

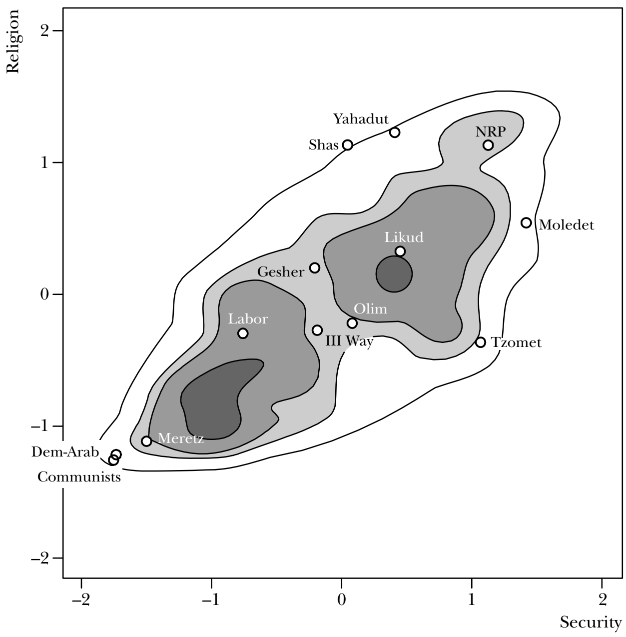


Fig. 1. Voter distribution and estimated party positions in the Knesset at the 1996 election

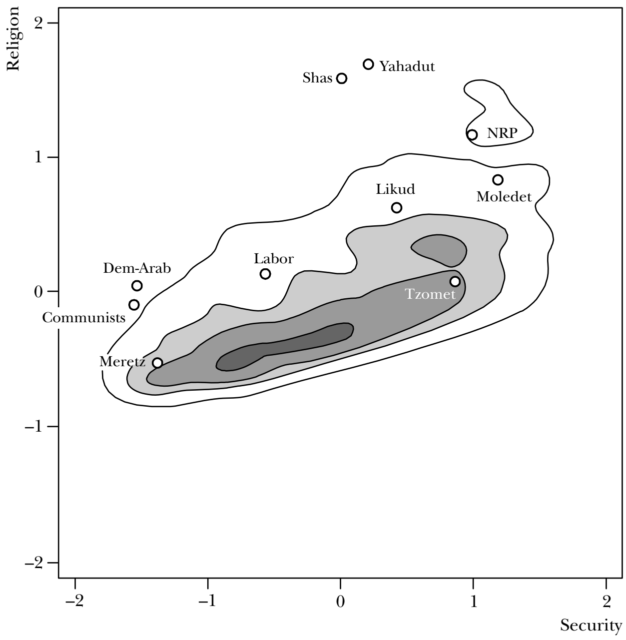


Fig. 2. Party positions and electoral distribution (at the 95%, 75%, 50% and 10% levels) in the Knesset at the 1992 election

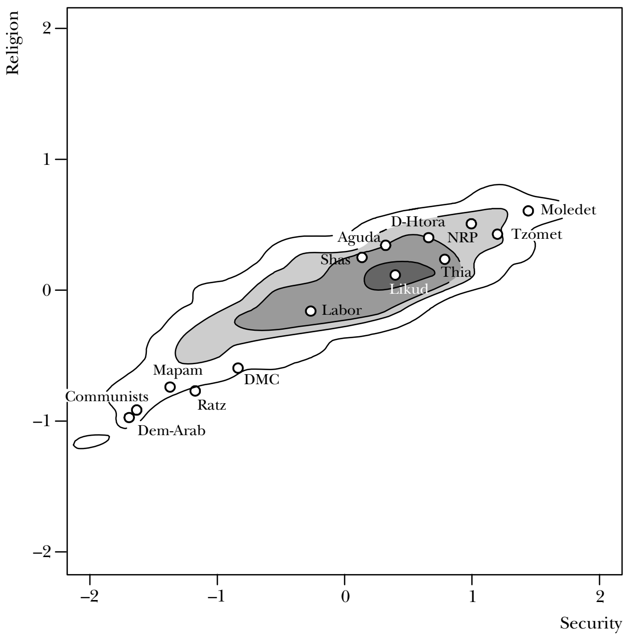


Fig. 3. Party positions and electoral distribution (at the 95%, 75%, 50% and 10% levels) in the Knesset at the 1988 election

Each respondent for the survey is characterized by a point in the resulting two-dimensional policy space, *X*. Thus the smoothed electoral distribution can be taken as an estimation of the underlying probability density function for the voter ideal points.

In these Figures, 'Security' refers to attitudes to peace initiatives. 'Religion' refers to the significance of religious considerations in government policy. The axes of the figures are oriented, so that 'left' on the security axis can be interpreted as supportive of negotiations with the PLO, while 'North' on the vertical or religious axis is indicative of a support for the importance of the Jewish faith in Israel (see Arien (1998), for discussion of these themes in the politics of Israel). Comparing Fig. 3 for 1988 with Fig. 1 for 1996 suggests that the covariance between the two factors has declined over time.

Since the competition between the two major parties, Labor and Likud, is pronounced, it is surprising that these parties do not move to the electoral mean (as suggested by the formal vote model) in order to increase vote and seat shares. The data on seats in the Knesset given in Table 1 suggest the vote share of the small Sephardic orthodox party, Shas, increased significantly between 1992 and 1996. As Fig.s 1 and 2 illustrate, however, there was no significant move by Shas to the electoral center. The inference is that the shifts of electoral support are the result of changes in party valence. To be more explicit, I contend that prior to an election each voter, i, forms a judgment about the relative capability of each party leader. Let λ_{ij} denote the weight given by voter i to party j in the voter's utility calculation. The voter utility is then given by the expression:

$$u_{ij}(x_i, z_j) = \lambda_{ij} - \beta \parallel x_i - z_j \parallel^2 + \theta_j^{\mathrm{T}} \eta_i. \tag{3}$$

However, these weights are subjective, and may well be influenced by idiosyncratic characteristics of voters and parties. For empirical analysis, I assume that $\lambda_{ij} = \lambda_j + \Delta_{ij}$, where Δ_{ij} is drawn at random from the Type I extreme value distribution, Ψ, for the variate ε_j, with expected value zero. The expected value, $Exp(\lambda_{ij})$, of λ_{ij} is λ_j, and so I write $\lambda_{ij} = \lambda_j + \varepsilon_j$, giving (2). Since I am mainly concerned with the voter's choice, I shall assume here that λ_j is exogenously determined. The details of the estimations of the parameters $[\beta; \lambda_j, j = 1, \ldots, p]$ and for the k by p matrix θ for 1996 are given in the Table 3.

Estimating the voter model requires information about sample voter behavior. It is assumed that data is available about voter intentions: this information is encoded, for each sample voter i by the vector $c_i = (c_{i1}, \ldots, c_{ip})$ where $c_{ij} = 1$ if and only if i intends to vote (or did indeed vote) for agent j. Given the data set $\{x_i, \eta_i, c_i\}_N$ for the sample N (of size n) and $\{z_j\}_P$, for the political agents, a set $\{\rho_i^*\}_N$ of stochastic variables is estimated. The first moment of ρ_i^* is the probability vector $\rho_i = \{\rho_{i1}, \ldots, \rho_{ip}\}$. Here ρ_{ij} is the probability that voter i chooses agent j.

There are standard procedures for estimating the empirical model. The technique is to choose estimators for the coefficients so that the estimated probability takes the form:

$$\overline{\rho}_{ij} = \Pr\left[\overline{u}_{ij}(x_i, z_j) > \overline{u}_{il}(x_i, z_l) \text{ for all } l \in P \setminus \{i\}\right]. \tag{4}$$

Here, $\overline{u}_{ij}$ is the j^{th} component of estimated latent utility function for i. The estimator for the choice is $\overline{c}_{ij} = 1$ if and only if $\overline{\rho}_{ij} > \overline{\rho}_{jl}$ for all $l \in P \setminus \{j\}$. The procedure minimizes the errors between the n by p matrix $[c]$ and the n by p estimated matrix $[\overline{c}]$. The vote share, $V_j^*(\mathbf{z})$, of agent i, given the vector z of strategies, is defined to be:

$$V_j^*(\mathbf{z}) = 1/n \Sigma_i \rho_{ij}^*(\mathbf{z}). \tag{5}$$

Note that since $V_j^*(\mathbf{z})$ is a stochastic variable, it is characterized by its first moment (its expectation), as well as higher moments (its standard variance, etc.). In the interpretation of the model I shall follow the usual assumption of the formal model and focus on the expectation $Exp(V_j^*(\mathbf{z}))$. The estimate of this expectation, denoted $E_j(\mathbf{z})$, is given by:

$$E_j(\mathbf{z}) = 1/n \Sigma_i \overline{\rho}_{ij}(\mathbf{z}). \tag{6}$$

Table 4. Bayes Factor $[\ln(B_{jk})]$ for Model *j* vis-à-vis Model *k* for the 1996 election

		Spatial MNP	Spatial MNL1	Spatial MNL2	Joint MNL
	Spatial MNP	n.a.	239^{***}	–17	–49
M_j	Spatial MNL1	–239	n.a.	–255	–288
	Spatial MNL2	17^{***}	255^{***}	n.a.	–33
	Joint MNL	49^{***}	288^{***}	33^{***}	n.a.

*** = very strong support for M_j

A virtue of using the general voting model is that the Bayes' factors (or differences in log likelihoods) can be used to determine which of various possible models is statistically superior (Quinn, Martin and Whitford 1999; Kass and Raftery 1995). I compared a variety of different MNL models against a pure MNP model for each election. The models were:

(i) MNP: a pure spatial multinomial probit model with $\beta \neq 0$ but $\theta \equiv 0$ and $\lambda = 0$.

(ii) MNLSD: a pure logit sociodemographic(SD) model, with $\beta = 0$, involving the component θ, based on respondent age, education, religious observance and origin (whether Sephardic, etc.).

(iii) MNL1: a pure multinomial logit spatial model with $\beta \neq 0$, but $\theta \equiv 0$ and $\lambda = 0$.

(iv) MNL2: a multinomial logit model with $\beta \neq 0$, $\theta \neq 0$ and $\lambda = 0$.

(v) Joint MNL: a multinomial logit model with $\beta \neq 0$, $\theta \neq 0$ and $\lambda \neq 0$.

The pure sociodemographic model MNLSD gave poor results and this model was not considered further. Table 4 give the comparisons for MNP, MNL1, MNL2 and Joint MNL for 1996

Note that the MNP model had no valence terms. Observe, for the 1996 election, the Bayes' factor for the comparison of the Joint MNL model with MNL1 was of order 288, so clearly sociodemographic variables add to predictive power. However, the valence constants add further to the power of the model. The spatial distance, as expected, exerts a very strong negative effect on the propensity of a voter to choose a given party. To illustrate, Table 3. shows that, in 1996, the β coefficient was estimated to be approximately 1.12. In short, Israeli voters cast ballots, to a very large extent, on the basis of the issue positions of the parties. This is true even after taking the demographic and religious factors into account. The coefficients on 'religious observation' for Shas and the NRP (both religious parties) were estimated to be 3.022 and 2.161 respectively. Consequently, a voter who is observant has a high probability of voting for one of these parties, but this probability appears to fall off rapidly the further is the voter's ideal position from the party position.

In each election, factors such as age, education, and religious observance play a role in determining voter choice. Obviously this suggests that some parties are more successful, among some groups in the electorate than would be implied by a simple estimation based only on policy positions.

These Tables indicates that the best model is the joint MNL that includes valence and the sociodemographic factors along with the spatial coefficient β. In particular, there is strong support, in all three elections, for the inclusion of valence. This model provides the best estimates of the vote shares of parties and predicts the vote choices of the individual voters remarkably well. Therefore this is clearly the model of choice to use as the best estimator for what I refer to as the stochastic electoral response function. Adding valence to the MNL model makes it superior to both MNL and MNP models without valence. Adding the sociological factors increases the statistical validity of the model.

Because the sociodemographic component of the model was assumed independent of party strategies, I could use the estimated parameters of the model to simulate party movement in order to increase the expected vote share of each party. 'Hill climbing' algorithms were used for this purpose. Such algorithms involve small changes in party position, and are therefore only capable of obtaining 'local' optima for each party. Consequently, a vector $\mathbf{z}^* = (z_1^*, \ldots, z_p^*)$ of party positions that results from such a search is what I call a 'local pure strategy Nash equilibrium' or LNE.

I now present the definition of local equilibria for the context of the empirical vote maximizing game defined by $E : X^p \to R^p$.

Definition 1 (i) A strategy vector $\mathbf{z}^* = (z_1^*, \ldots, z_{j-1}^*, z_j^*, z_{j+1}^*, \ldots, z_p^*) \in X^p$ is a *local strict Nash equilibrium* (LSNE) for the profile function $E : X^p \to R^p$ iff, for each agent $j \in P$, there exists a neighborhood X_j of z_j^* in X such that:

$$E_j(z_1^*, \ldots, z_{j-1}^*, z_j^*, z_{j+1}^*, \ldots, z_p^*) > E_j(z_1^*, \ldots, z_j, \ldots, z_p^*) \text{ for all } z_j \in X_j - \{z_j^*\}.$$

(ii) A strategy vector $\mathbf{z}^* = (z_1^*, \ldots, z_{j-1}^*, z_j^*, z_{j+1}^*, \ldots, z_p^*)$ is a *local weak Nash equilibrium* (LNE) for E iff, for each agent j, there exists a neighborhood X_j of z_j^* in X such that:

$$E_j(z_1^*, \ldots, z_{j-1}^*, z_j^*, z_{j+1}^*, \ldots, z_p^*) \geq E_j(z_1^*, \ldots, z_j, \ldots, z_p^*) \text{ for all } z_j \in X_j.$$

(iii) A strategy vector $\mathbf{z}^* = (z_1^*, \ldots, z_{j-1}^*, z_j^*, z_{j+1}^*, \ldots, z_p^*)$ is a *strict*, respectively, *weak, pure strategy Nash equilibrium* (PSNE, respectively, PNE) for E iff X_j can be replaced by X in (i), (ii) respectively.

(iv) The strategy z_j^* is termed a 'local strict best response,' a 'local weak best response,' a 'global weak best response,' a 'global strict best response,' respectively to $\mathbf{z}_{-j}^* = (z_1^*, \ldots; z_{j-1}^*; z_{j+1}^*; \ldots, z_p^*)$.

In these definitions 'weak' refers to the condition that z_j^* is no worse than any other strategy. Clearly, a LSNE must be a LNE, and a PNE must be a LNE. Obviously a LNE need not be a PNE. If a necessary condition for LNE fails, then so does the necessary condition for PNE. One condition that is sufficient to guarantee that a LNE is a PNE for the electoral game is *concavity* of the vote functions.

Definition 2 The profile is *concave* $E : X^p \to R^p$ iff for each *j*, and any real α and $x, y \in X$, then $E_j(\alpha x + (1-\alpha)y) \geq \alpha E_j(x) + (1-\alpha)E_j(y)$.

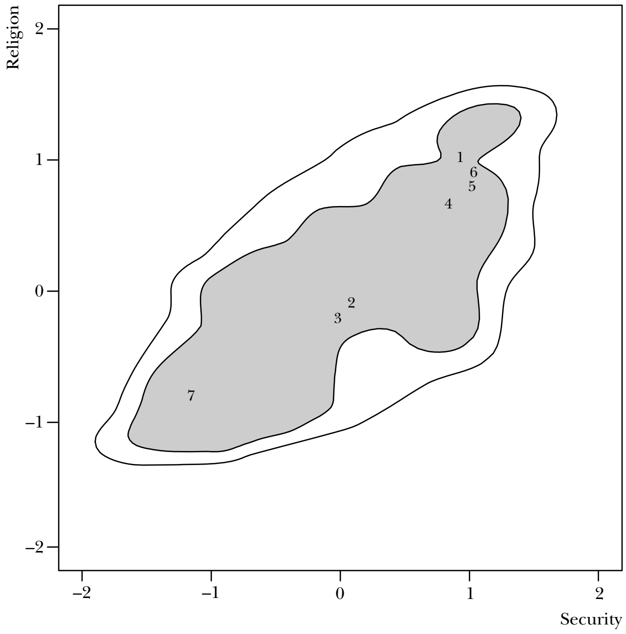
Concavity of the payoff functions $\{E_j\}$ in the *j*-th strategy z_j, together with continuity in z_j and compactness and convexity of X is sufficient for existence of PNE (Banks and Duggan 2005). The much weaker condition that $\mathbf{z}^*$ be a LNE is *local concavity,* namely that concavity holds in a neighborhood of $\mathbf{z}^*$. We use this in the following section. There, I discuss the 'mean voter theorem' of the formal model. As mentioned above, this theorem asserts that the vector $\mathbf{z}^* = (x^*,\ldots,x^*)$ (where x^* is the mean of the distribution of voter ideal points) is a PNE for the vote maximizing electoral game (Hinich 1977; Enelow and Hinich 1984; Lin, Enelow, and Dorussen 1999). I call $(x^*,\ldots,x^*)$ the *joint electoral mean.* Since the electoral distribution can be readily normalized, so $x^* = 0$, I shall also use the term *joint electoral origin.* I used a hill climbing algorithm to determine the LSNE of the empirical vote models for the three elections.

The simulation of the empirical models found five distinct LNE for the 1996 election in Israel. A representative LNE is given in Fig. 4. Notice that the locations of the two high valence parties, Labor and Likud, in Fig. 1 closely match their simulated positions in Fig. 4. Obviously, none of the estimated equilibrium vectors in Fig. 4 correspond to the convergent situation at the electoral mean. Figs. 5 and 6 give representative LNE for 1992 and 1988. Before I begin the theoretical discussion of the results just presented, several preliminary conclusions appear to be of interest.

(i) The empirical MNL model and the formal model, discussed in the next section are both based on the extreme value distribution Ψ and are mutually compatible.

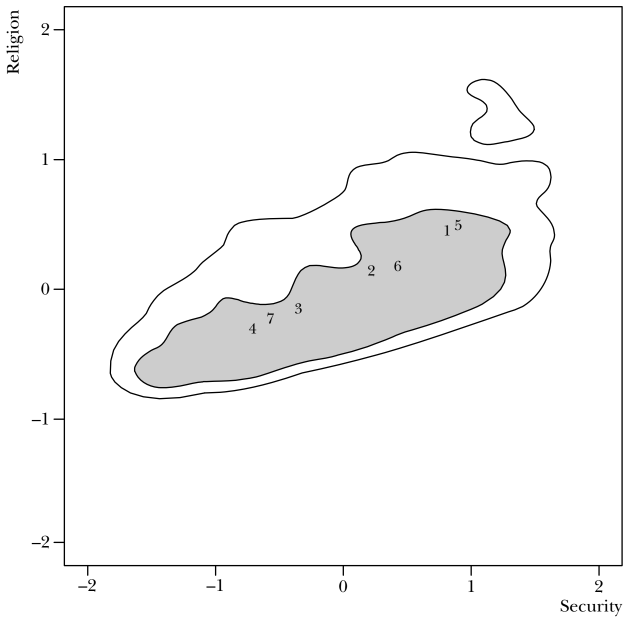
(ii) The set of LNE obtained by simulation of the empirical model must contain any PNE for this model (if any exist). Since no LSNE was found at the joint mean position, it follows that the mean voter theorem is invalid, given the estimated parameter values of the empirical model. This conclusion is not susceptible to any counter-argument that the parties may have utilized evaluation functions other than expected vote shares, because only vote share maximization was allowed to count in the 'hill climbing' algorithm used to generate the LNE.

(iii) A comparison of Figs. 1, 2 and 3 with the simulation figures 4, 5 and 6 makes it clear that there are marked similarities between estimated and simulated positions. This is most obvious for the high valence parties, Labor and Likud, but also for the low valence party Meretz. This suggests that the expected vote share functions $\{E_j\}$ is a close proxy to the actual, but unknown, utility functions $\{U_j\}$, deployed by the party leaders.



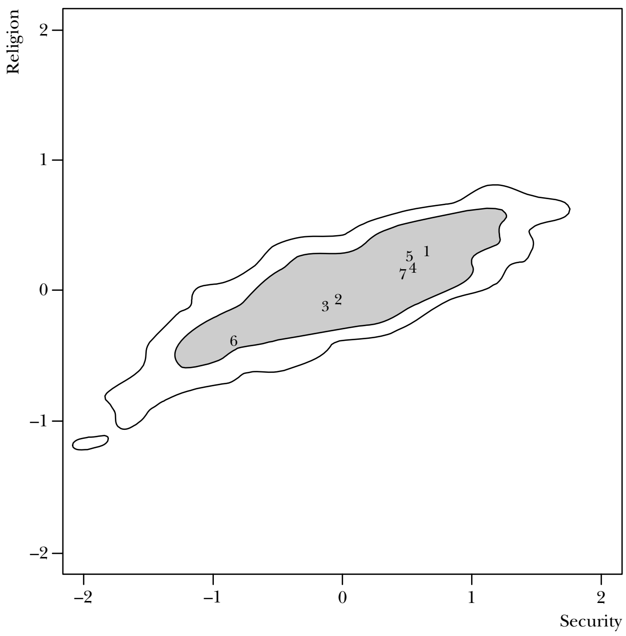
Key 1=Shas, 2=Likud, 3=Labor, 4=NRP, 5=Molodet, 6= Third Way, 7=Meretz.

Fig. 4. Simulated Local Nash Equilibrium for Israel in 1996



Key 1=Shas, 2=Likud, 3=Labor, 4=Meretz, 5=NRP, 6=Molodet, 7=Tzomet.

Fig. 5. A representative Local Nash Equilibrium in the Knesset, 1992 election



Key 1=Shas, 2=Likud, 3=Labor, 4=Tzomet, 5=NRP, 6=Ratz, 7=Thia.

Fig. 6. A representative Local Nash Equilibrium in the Knesset, 1988 election

(iv) Although the equilibrium notion of LSNE that I deploy is not utilized in the game theoretic literature, it has a number of virtues. In particular, the simulation performed here shows that local equilibria do indeed exist. Other formal results (Schofield 2001) show that such equilibria typically exist, as long as the game form is 'smooth' Moreover, the definition of $\{E_j\}$ presented above makes it obvious that the vote maximization game considered here is indeed smooth. On the other hand existence for PNE is problematic when concavity fails.

(v) Although the 'local' equilibrium concept is indeed 'local,' there is no formal reason why each of the various LNE that I obtain should be, in fact, 'close' to one another. It is noticeable in Figs. 4, 5 and 6 that the LNE for each election are approximately permutations of one another, with low valence parties strung along what I shall call the electoral principal axis.

In the following section, I examine the formal vote model in order to determine why the mean voter theorem appears to be invalid for the estimated model of Israel. The formal result will explain why low valence parties in the simulations are far from the electoral mean, and why all parties lie on a single electoral axis.

4. The Formal Model of Elections

Given the vector of agent policy positions. $\mathbf{z} = (z_1, \ldots, z_p) \in X^p$,each voter, i, is described by a vector $\mathbf{u}_i(x_i, \mathbf{z}) = (u_{i1}(x_i, z_1), \ldots, u_{ip}(x_i, z_p))$, where:

$$u_{ij}(x_i, z_j) = u^*_{ij}(x_i, z_j) + e_j$$
$$\text{and } u^*_{ij}(x_i, z_j) = \lambda_j - \beta \,||\, x_i \text{ - } z_j \,||^2 .$$

Here, $u^*_{ij}(x_i, z_j)$ is the 'observable' utility for i associated with party j The valence λ_j of agent j is exogenously determined. The terms $\{\varepsilon_j\}$ are the stochastic errors,whose cumulative distibution is denoted by Ψ.

Again, the probability that a voter i chooses party j is:

$$\rho_{ij}(\mathbf{z}) = \Pr\left[[u_{ij}(x_i, z_j) > u_{il}(x_i, z_l)], \text{ for all } l \neq j \right].$$

Here Pr stands for the probability operator associated with Ψ. The expected vote share of agent j is:

$$V_j(\mathbf{z}) = \frac{1}{n} \sum_{i \in N} \rho_{ij}(\mathbf{z}).$$

In the vote model it is assumed that each agent j chooses z_j to maximize V_j , conditional on $\mathbf{z}_{-j} = (z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_p)$.

The theorem presented here assumes that the exogeneous valences are given by the vector $\lambda = (\lambda_p, \lambda_{p-1}, \ldots, \lambda_2, \lambda_1)$ and the valances are ranked $\lambda_p \geq \lambda_{p-1} \geq \ldots \geq \lambda_2 \geq \lambda_1$. The model is denoted $M(\lambda, \beta, \Psi)$.

In this model it is natural to regard λ_j as the 'average' weight given by a member of the electorate to the perceived competence or quality of agent j. The 'weight' will in fact vary throughout the electorate, in a way which is described by the stochastic distribution. In these models, the C^2-differentiability of the cumulative distribution, Ψ, is usually assumed, so that the individual probability functions $\{\rho_{ij}\}$ will be C^2-differentiable in the strategies $\{z_j\}$. Thus, the vote share functions will also be C^2-differentiable and Hessians can be calculated.

Let $x^* = (1/n)\Sigma_i x_i$. Then the mean voter theorem for the stochastic model asserts that the 'joint mean vector' $\mathbf{z}^*_0 = (x^*, \ldots, x^*)$ is a 'pure strategy Nash equilibrium (PNE). Adapting Definition 1, we say a strategy vector $\mathbf{z}^*$ is a LSNE for the formal model iff, for each j, z^*_j is a critical point of the vote function $V_j(z^*_1, \ldots, z^*_{j-1}, -z^*_{j+1}, \ldots, z^*_p)$ and the eigenvalues of the Hessian of this function (with respect to z_j), are negative. The definitions of LNE, PSNE and PNE for the profile $V : X^p \to R^p$ for the formal model follow Definition 1 above.

The theorem below gives the necessary and sufficient conditions for the joint mean vector $\mathbf{z}^*_0$ to be an LSNE. A corollary of the theorem shows, in situations where the valences differ, that the necessary condition is likely to fail. In dimension w, the theorem can be used to show that, for $\mathbf{z}^*_0$ to be an

LSNE, the necessary condition is that a 'convergence coefficient,' defined in terms of the parameters of the model, must be strictly bounded above by w. When this condition fails, then the joint mean vector $\mathbf{z}_0^*$ cannot be a LNE and therefore cannot be a PNE. Of course, even if the sufficient condition is satisfied, and $\mathbf{z}_0^* = (x^*, \dots, x^*)$ is an LSNE, it need not be a PNE.

As with the empirical analysis, I assume a Type I extreme value distribution (Train 2003) for the errors.

The cumulative distribution, Ψ, takes the closed form:

$$\Psi(h) = \exp[-\exp[-h]]$$

with variance is $\frac{1}{6}\pi^2$. It readily follows (Train 2003: 79) for the choice model given above that, for each i,

$$\rho_{ij}(\mathbf{z}) = \frac{\exp\left[u_{ij}^*(x_i, z_j)\right]}{\sum_{k=1}^{p} \exp\left[u_{ik}^*(x_i, z_k)\right]}.$$

This implies that the model satisfies the independence of irrelevant alternative property (IIA) namely that for each voter i, and for each pair, j,k, the ratio $\rho_{ij}(\mathbf{z})/\rho_{ik}(\mathbf{z})$ is independent of a third candidate l.

To state the theorem, I first transform coordinates so that in the new coordinates, $x^* = 0$. I shall refer to $\mathbf{z}_0^* = (0, \dots, 0)$ as the *joint origin* in this new coordinate system. Whether the joint origin is an equilibrium depends on the distribution of voter ideal points. These are encoded in the voter variance/covariance matrix. I first define this, and then use it to characterize the vote share Hessians.

Definition 3 *The voter variance-covariance matrix* (∇^*). To characterize the variation in voter preferences, I represent in a simple form the variance covariance matrix, ∇^* of the distribution of voter ideal points. Let X have dimension w and be endowed with a system of coordinate axes $(1, \dots, s, t, \dots, w)$. For each coordinate axis let $\xi_t = (x_{1t}, x_{2t}, \dots, x_{nt})$ be the vector of the t^{th} coordinates of the set of n voter ideal points. I use (ξ_s, ξ_t) to denote scalar product.

The symmetric $w \times w$ voter covariation or data matrix ∇ is then defined to be:

$$\nabla = \begin{pmatrix} (\xi_1, \xi_1) & & & (\xi_1, \xi_w) \\ & (\xi_s, \xi_s) & & \\ & & (\xi_t, \xi_t) & \\ (\xi_w, \xi_1) & & & (\xi_w, \xi_w) \end{pmatrix}.$$

The covariance matrix is $\nabla^* = \frac{1}{n}\nabla$. I write $v_s^2 = \frac{1}{n}(\xi_s, \xi_s)$ for the electoral variance on the s^{th} axis and

$$v^2 = \sum_{r=1}^{w} v_r^2 = \frac{1}{n}\sum_{r=1}^{w} (\xi_r, \xi_r) = trace(\nabla^*)$$

for the total electoral variance. The normalized covariance between the s-th and t-th axes is $(v_s, v_t) = \frac{1}{n}(\xi_s, \xi_t)$.

The formal model is denoted $M(\lambda, \beta; \Psi, \nabla^*)$, though I shall usually suppress the reference to ∇^*.

Definition 4 *The Convergence Coefficient of the model* $M(\lambda, \beta; \Psi)$. (i) At the vector $\mathbf{z}_0 = (0,..,0)$ the probability $\rho_k(\mathbf{z}_0)$ that i votes for party, k, is

$$\rho_k = \left[1 + \sum\nolimits_{j \neq k} \exp[\lambda_j - \lambda_k]\right]^{-1}.$$

(ii) The coefficient A_k for party k is

$$A_k = \beta(1 - 2\rho_k).$$

(iii) The characteristic matrix for party k at $\mathbf{z}_0$ is

$$C_k = \left[2[A_k](\nabla^*) - I\right].$$

where I is the w by w identity matrix.

(iv) The *convergence coefficient of the model* $M(\lambda, \beta; \Psi)$ is

$$c(\lambda, \beta; \Psi) = 2\beta[1 - 2\rho_1]v^2 = 2A_1 v^2.$$

The definition of ρ_k follows directly from the definition of the extreme value distribution. Obviously, if all valences are identical, then $\rho_1 = \frac{1}{p}$, as expected. The effect of increasing λ_j, for $j \neq 1$, is clearly to decrease ρ_1, and therefore to increase A_1, and thus $c(\lambda, \beta; \Psi)$.

Theorem 3.1 *The electoral theorem* (Schofield 2006*a*, 2007*a,b*) The condition for the joint origin be a LSNE in the model $M(\lambda, \beta; \Psi)$ is that the characteristic matrix

$$C_1 = \left[2A_1 \nabla^* - I\right]$$

of the party 1, with lowest valence, has negative eigenvalues.

As usual, conditions on C_1 for the eigenvalues to be negative depend on the trace, trace(C_1), and determinant, $\det(C_1)$, of C_1 (see Schofield (2003) for these standard results). These depend on the value of A_1 and on the electoral covariance matrix, ∇^*. Using the determinant of C_1, it is easy to show, in two dimensions, that $2A_1 v^2 < 1$ is a sufficient condition for the eigenvalues to be negative. In terms of the 'convergence coefficient'

$c(\lambda,\beta;\Psi)$ I can write this as $c(\lambda,\beta;\Psi)<1$. In a policy space of dimension w, the necessary condition on C_1, induced from the condition on the Hessian of V_1, is that $c(\lambda,\beta;\Psi)\leq w$. If this necessary condition for V_1 fails, then $\mathbf{z}_0$ can be a neither a LNE nor a LSNE.

Ceteris paribus, a LNE at the joint origin is 'less likely' the greater are the parameters β, $\lambda_p-\lambda_1$ and v^2.

The Theorem gives the following Corollaries.

Corollary 1 Suppose the dimension of X is w. Then, in the model $M(\lambda,\beta;\Psi)$, the necessary condition for the joint origin to be a LNE is that $c(\lambda,\beta;\Psi)$ be no greater than w.

Notice that the case with two parties with equal valence immediately gives a situation with $2\beta[1-2\rho_1]v^2=0$, irrespective of the other parameters. However, if $\lambda_2>\lambda_1$, then the joint origin may fail to be a LNE if βv^2 is sufficiently large. Note also that for the multiparty case ρ_1 is a decreasing function of $(\lambda_p-\lambda_1)$ so the necessary condition is more difficult to satisfy as $(\lambda_p-\lambda_1)$ increases.

Corollary 2 In the two dimensional case, the sufficient condition for the joint origin to be a LSNE is that $c(\lambda,\beta;\Psi)$ be strictly less than 1. Furthermore if $v_t^2=\frac{1}{n}(\xi_t,\xi_t)$ are the electoral variances on the two axes $t=1,2$, then the two eigenvalues of C_1 are:

$$a_1=(A_1)\left\{\left[v_1^2+v_2^2\right]+\left[[v_1^2-v_2^2]^2+4(v_1,v_2)^2\right]^{\frac{1}{2}}\right\}-1,$$
$$a_2=(A_1)\left\{\left[v_1^2+v_2^2\right]-\left[[v_1^2-v_2^2]^2+4(v_1,v_2)^2\right]^{\frac{1}{2}}\right\}-1.$$

When the covariance term $(v_1,v_2)=0$, then the eigenvalues are obviously $a_t=A_1v_t^2, t=1,2$. The more interesting case is when the covariance (v_1,v_2) is significant. By a transformation of coordinates, I can choose v_t, v_s to be the eigenvectors of the Hessian matrix for agent 1, and let these be these new 'principal components' of the electoral covariance matrix. If $v_t^2\leq v_s^2$ then the s-th coordinate can be termed 'the principal electoral axis.'

5. Empirical Analysis for Israel

Consider first the election for the Israel Knesset in 1996. Using the formal analysis, I can readily show that one of the eigenvalues of the low valence party, the NRP, is positive. Indeed it is obvious that there is a principal component of the electoral distribution, and this axis is the eigenspace of the positive eigenvalue. It follows that low valence parties should then position themselves on this eigenspace as illustrated in the simulation given below in Fig. 4.

In 1996, the lowest valence party was the NRP with valence −4.52. The spatial coefficient is $\beta = 1.12$, so for the extreme value model $M(\Psi)$ I compute $\rho_{NRP} \simeq 0$ and $A_{NRP} = 1.12$. Now $v_1^2 = 1.0, v_2^2 = 0.732$ and $(v_1, v_2) = 0.591$ (see Table 2). Thus

$$\begin{aligned}
\rho_{NRP} &\simeq \frac{1}{1 + e^{4.15+4.52} + e^{3.14+4.52}} \simeq 0, \\
A_{NRP} &= \beta = 1.12, \\
C_{NRP} &= 2(1.12)\begin{pmatrix} 1.0 & 0.591 \\ 0.591 & 0.732 \end{pmatrix} - I = \begin{pmatrix} 1.24 & 1.32 \\ 1.32 & 0.64 \end{pmatrix}, \\
c(\Psi) &= 3.88.
\end{aligned}$$

Then the eigenvalues are 2.28 and −0.40, giving a saddlepoint, and a value for the convergence coefficient of 3.88. The major eigenvector for the NRP is $(1.0, 0.8)$, and along this axis the NRP vote share function increases as the party moves away from the origin. The minor, perpendicular axis is given by the vector $(1.0, -1.25)$ and on this axis the NRP vote share decreases. Fig. 4 gives one of the local equilibria in 1996, obtained by simulation of the model. The figure makes it clear that the vote maximizing positions lie on the principal axis through the origin and the point $(1.0, 0.8)$. In all, five different LSNE were located. However, in all the equilibria, the two high valence parties, Labor and Likud, were located at positions similar to those given in Fig. 4. The only difference between the various equilibria were that the positions of the low valence parties were perturbations of each other.

I next analyze the situation for 1992, by computing the eigenvalues for the Type I extreme value distribution, Ψ. From the empirical model, I obtain $\lambda_{shas} = -4.67, \lambda_{likud} = 2.73, \lambda_{labour} = 0.91, \beta = 1.25$. When all parties are at the origin, then the probability that a voter chooses Shas is

$$\rho_{shas} \simeq \frac{1}{1 + e^{2.73+4.67} + e^{0.91+4.67}} \simeq 0.$$

Thus,

$$\begin{aligned}
A_{shas} &= \beta = 1.25, \\
C_{shas} &= 2(1.25)\begin{pmatrix} 1.0 & 0.453 \\ 0.453 & 0.435 \end{pmatrix} - I = \begin{pmatrix} 1.5 & 1.13 \\ 1.13 & 0.08 \end{pmatrix}, \\
c(\Psi) &= 3.6.
\end{aligned}$$

Then the two eigenvalues for Shas can be calculated to be $+2.12$ and -0.52 with a convergence coefficient for the model of 3.6. Thus I find that the origin is a saddlepoint for the Shas Hessian. The eigenvector for the large, positive eigenvalue is the vector $(1.0, 0.55)$. Again, this vector coincides with the principal electoral axis. The eigenvector for the negative eigenvalue is perpendicular to the principal axis. To maximize vote share, Shas should adjust its position but only on the principal axis. This is exactly what the simulation found. Notice that the probability of voting for Labor is $[1+e^{1.82}]^{-1} = 0.14$, and $A_{labour} = 0.9$, so even Labor will have a positive eigenvalue at the origin. Clearly, if Likud occupies the mean voter position, then Labor as well as all low valence parties would find this same position to be a saddlepoint. In seeking local maxima of vote shares all parties other than Likud should vacate the electoral center. Then, however, the first order condition for Likud to occupy the electoral center would not be satisfied. Even though Likud's vote share will be little affected by the other parties, it too should move from the center. This analysis predicts that the lower the party's valence, the further will its equilibrium position be from the electoral mean. This is illustrated in Fig. 5.

Calculation for the model $M(\Psi)$ for 1988 gives eigenvalues for Shas of $+2.0$ and -0.83 with a convergence coefficient of 3.16, and a principal axis through $(1.0, 0.5)$. Again, vote maximizing behavior by Shas should oblige it to stay strictly to the principal electoral axis. The simulated vote maximizing party positions indicated that there was no deviation by parties off the principle axis or eigenspace associated with the positive eigenvalue.

Thus the simulation was compatible with the predictions of the formal model based on the extreme value distribution. All parties were able to increase vote shares by moving away from the origin, along the principal axis, as determined by the large, positive principal eigenvalue. In particular, the simulation confirms the logic of the formal analysis. Low valence parties, such as the NRP and Shas, in order to maximize vote shares must move far from the electoral center. Their optimal positions will lie either in the 'north east' quadrant or the 'south west' quadrant. The vote maximizing model, without any additional information, cannot determine which way the low valence parties should move. As noted above, the simulations of the empirical models found multiple LSNE essentially differing only in permutations of the low valence party positions.

In contrast, since the valence difference between Labor and Likud was relatively low in all three elections, their optimal positions would be relatively close to, but not identical to, the electoral mean. The simulation figures for all three elections are also compatible with this theoretical inference. The figures also suggest that every party, in local equilibrium, should adopt a position that maintained a minimum distance from every other party. The formal analysis, as well as the simulation exercise, suggests that this minimum distance depends on the valences of the neighboring parties. Intuitively it is clear that once the low valence parties vacate the origin, then high valence parties, like Likud and Labor will position themselves almost symmetrically about the origin, and along the major axis. It should be noted that the positions of Labor and Likud, particularly, closely match their positions in the simulated vote maximizing equilibria.

Clearly, the configuration of equilibrium party positions will fluctuate as the valences of the large parties change in response to exogenous shocks. The logic of the model remains valid however, since the low valence parties will be obliged to adopt relatively 'radical' positions in order to maximize their vote shares.

The correlation between the two electoral axes was much higher in 1988 $(r^2 = 0.70)$ than in 1992 or 1996 (when $r^2 \simeq 0.47$). It is worth observing that as r^2 falls from 1988 to 1996, a counter-clockwise rotation of the principal axis that can be observed. This can be seen in the change from the eigenvalue $(1.0, 0.5)$ in 1988, to $(1.0, 0.55)$ in 1992 and then to $(1.0, 0.8)$ in 1996. Notice also that the total electoral variance increased from 1988 to 1992 and again from 1992 to 1996. Indeed, in 1996, Fig. 1 indicates that there is evidence of bifurcation in the electoral distribution in 1996.

In comparing Fig. 1, of the estimated party positions, and Fig. 4, of simulated equilibrium positions, there is a notable disparity particularly in the position of Shas. In 1996, Shas was pivotal between Labor and Likud, in the sense that to form a winning coalition government, either of the two larger parties required the support of Shas. It is obvious that the location of Shas in Fig. 1 suggests that it was able to bargain effectively over policy, and presumably perquisites. Indeed, it is plausible that the leader of Shas was aware of this situation, and incorporated this awareness in the utility function of the party.

The relationship between the empirical work and the formal model, together with the possibility of strategic reasoning of this kind, suggests that the true but unknown utility functions for the political game are perturbations of the expected vote share functions. These perturbations may be modelled by taking into account the post election coalition possibilities, as well as the effect of activist support for the party.

6. Concluding Remarks

The purpose of this paper has been to argue that the 'centripetal' tendency of political strategy, inferred from the spatial voting model, is contradicted by empirical evidence. It is suggested, instead, that a valence model can be utilized both to predict individual vote choice, and to interpret the political game. The necessary condition for local Nash equilibrium at the electoral mean also gives a necessary condition for vote maximizing Nash equilibrium at the mean. The validity of the 'mean voter theorem' depends essentially on a limit to the electoral variance, as well as bounds on the variation of valence between the parties. The evidence from Israel, with a highly proportional electoral system, suggests that the mean voter theorem can generally be expected to fail.

Empirical models for other polities can be used to examine the variation between predicted positions and estimated positions, and these differences would, in principle, allow us to infer the true nature of the utility functions of party leaders. It is plausible that party calculations are affected by considerations of activist support, as suggested by Aldrich (1983*a,b*). In principle, the model of exogenous valence proposed here can be extended by allowing valence to be affected by activist support in the electorate. The degree to which parties will locate themselves far from the center will depend on the nature of the activist calculus. Under proportional electoral rules, it is likely that there will be many potential activist coalitions, and their interaction will determine the degree of political fragmentation, as well as the configuration of small radical parties. Refining the formal model in this way may suggest how the simple vote maximization model may be extended to provide a better account of party position taking.

The relationship between the empirical work and the formal model, together with the possibility of strategic reasoning of this kind, suggests that the true but unknown utility functions for the political game are perturbations of the expected vote share functions. These perturbations may be modelled by taking into account the post election coalition possibilities, as well as the effect of activist support for the party.

Acknowledgements

This paper is based on research supported by NSF Grant SES 024173. Empirical and computational work mentioned here was undertaken in collaboration with Itai Sened. The tables and figures are taken from Schofield and Sened (2006) with pemission of Cambridge University Press.

References

Adams, J. (1999*a*) Multiparty Spatial Competition with Probabilistic Voting, *Public Choice* 99: 259–274.

Adams, J. (1999*b*) Policy Divergence in Multicandidate Probabilistic Spatial Voting, *Public Choice* 100: 103–122.

Adams, J. (2001) *Party Competition and Responsible Party Government*, University of Michigan Press.

Adams, J. and Merrill, S. (1999) Modeling Party Strategies and Policy Representation in Multiparty Elections: Why are Strategies so Extreme? *American Journal of Political Science* 43: 765–791.

Adams, J. and Merrill, S. (2000) Spatial Models of Candidate Competition and the 1988 French Presidential Election: Are Presidential Candidates Vote Maximizers? *Journal of Politics* 62: 729–756.

Adams, J. and Merrill, S. (2001) Computing Nash Equilibria in Probabilistic Multiparty Spatial Models with Non-policy Components, *Political Analysis* 9: 347–361.

Adams, J. and Merrill, S. (2005) Policy Seeking Parties in a Parliamentary Democracy with Proportional Representation: A Valence-Uncertainty Model, mimeo. University of California at Davis.

Adams, J. Merrill, S. and Grofman, B. (2005) *A Unified Theory of Party Competition*, Cambridge University Press.

Aldrich, J. (1983*a*) A Spatial Model with Party Activists: Implications for Electoral Dynamics, *Public Choice* 41: 63–100.

Aldrich, John. (1983*b*) A Downsian Spatial Model with Party Activists, *American Political Science Review* 77: 974–990.

Aldrich, J. (1995) *Why Parties?* Chicago University Press.

Aldrich, J. and McGinnis, M. (1989) A Model of Party Constraints on Optimal Candidate Positions, *Mathematical and Computer Modelling* 12: 437–450.

Alvarez, M. and Nagler, J. (1998) When Politics and Models Collide: Estimating Models of Multi-candidate Elections, *American Journal of Political Science* 42: 55–96.

Alvarez, M., Nagler, J. and Bowler, S. (2000) Issues, Economics, and the Dynamics of Multiparty Elections: The British 1987 General Election, *American Political Science Review* 94: 131–150.

Alvarez, M., Nagler, J. and Willette, J. (2000) Measuring the Relative Impact of Issues and the Economy in Democratic Elections, *Electoral Studies* 19: 237–253

Ansolabehere, S. and Snyder, J. (2000) Valence Politics and Equilibrium in Spatial Election Models, *Public Choice* 103: 327–336.

Aragones, E. and Palfrey, T. (2002) Mixed Equilibrium in a Downsian Model with a favored Candidate, *Journal of Economic Theory* 103: 131–161.

Aragones, E. and Palfrey, T. (2005) Spatial Competition between Two Candidates of Different Quality: The Effects of Candidate Ideology and Private Information, in D. Austen-Smith, and J. Duggan (eds) *Social Choice and Strategic Decisions*, Springer.

Arian, A. (1998) *The Second Republic: Politics in Israel*, Chatham House.

Arian, A. and Shamir, M. (1990) *The Election in Israel: 1988*, SUNY Press.

Arian, A. and Shamir, M. (1995) *The Election in Israel: 1992*, SUNY Press.

Arian, A. and Shamir, M. (1999) *The Election in Israel: 1996*, SUNY Press.

Austen-Smith, D. and Banks, J. (1998) Social Choice Theory, Game Theory and Positive Political Theory, *Annual Review of Political Science* 1: 259–287.

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

Banks, J. and Duggan, J. (2005) The Theory of Probabilistic Voting in the Spatial Model of Elections, in D. Austen-Smith and J. Duggan (eds) *Social Choice and Strategic Decisions*, Springer.

Budge, I., Robertson, D. and Hearl, D. (eds) (1987) *Ideology, Strategy and Party Change: A Spatial Analysis of Post-War election Programmes in Nineteen Democracies*, Cambridge University Press.

Calvert, R. (1985) Robustness of the Multidimensional Voting Model: Candidates, Motivations, Uncertainty and Convergence, *American Journal of Political Science* 29: 69–85.

Chib, S. and Greenberg, E. (1996) Markov Chain Monte Carlo Simulation Methods in Econometrics, *Econometric Theory* 12: 409–431.

Clarke, H., Stewart, M. and Whiteley, P. (1997) Tory Trends, Party Identification and the Dynamics of Conservative Support since 1992, *British Journal of Political Science* 26: 299–318.

Clarke, H., Stewart, M. and Whiteley, P. (1998) New Models for New Labour: The Political Economy of Labour Support, *American Political Science Review* 92: 559–575.

Clarke, H. et al. (2004) *Political Choice in Britain*, Oxford University Press.

Condorcet, N. (1795) *Essai sur lapplication de lanalyse a la probabilite des decisions rendus a la pluralite des voix*, Imprimerie Royale.

Coughlin, P. (1992) P*robabilistic Voting Theory*, Cambridge University Press.

Cox, G. (1990) Centripetal and Centrifugal Incentives in Alternative Voting Institutions, *American Journal of Political Science* 34: 903–935.

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

Daalder, H. (1984) In Search of the Center of European Party Systems, *American Political Science Review* 78: 92–109.

Dow, J.K. and Endersby, J.W. (2004) Multinomial Probit and Multinomial Logit: A Comparison of Choice Models for Voting Research, *Electoral Studies* 23: 107–122.

Downs, A. (1957) *An Economic Theory of Democracy*, Harper and Row.

Duverger, M. (1954) *Political Parties: Their Organization and Activity in the Modern State*, Wiley.

Enelow, J. and Hinich, M. (1982) Nonspatial Candidate Characteristics and Electoral Competition, *Journal of Politics* 44: 115–130.

Enelow, J. and Hinich, M. (1984*a*) Probabilistic Voting and the Importance of Centrist Ideologies in Democratic Elections, *Journal of Politics* 46: 459–478.

Enelow, J. and Hinich, M. (1984*b*) *The Spatial Theory of Voting*, Cambridge University Press.

Enelow, J. and Hinich, M. (1989*a*) A General Probabilistic Spatial Theory of Elections, *Public Choice* 61: 101–114.

Enelow, J. and Hinich, M. (1989*b*) The Location of American Presidential Candidates: An Empirical Test of a New Spatial Model of Elections, *Mathematical and Computer Modelling* 12: 461–470.

Glasgow, G. and Alvarez, M. (2003) Voting Behavior and the Electoral Context of Government Formation, *Electoral Studies* 24: 245–264.

Groseclose, T. (2001) A Model of Candidate Location when one Candidate has a Valance Advantage, *American Journal of Political Science* 45: 862–886.

Hinich, M. (1977) Equilibrium in Spatial Voting: The Median Voter Result is an Artifact, *Journal of Economic Theory* 16: 208–219.

ISEIUM (1983) European Elections Study: European Political Parties Middle Level Elites, Europa Institut, Mannheim.

Kass, R. and Raftery, A. (1995) Bayes Factors, *Journal of the American Statistical Association* 90: 773–795.

Laver, M. and Schofield, N. (1990) *Multiparty Government*, Oxford University Press.

Laver, M. and Shepsle, K. (1996) *Making and Breaking Governments*, Cambridge University Press.

Lin, T-M., Enelow, J. and Dorussen, H. (1999) *Equilibrium in Multicandidate Probabilistic Spatial Voting*, Public Choice 98: 59–82.

Madison, J. (1787) The Federalist No.10, in J. Rakove (ed.) *James Madison: Writings*, The Library of America.

McKelvey, R. and Patty, J. (2006) A Theory of Voting in Large Elections, *Games and Economic Behavior* 57: 155–180.

McLennan, A. (1998) Consequences of the Condorcet Jury Theorem for Beneficial

Information Aggregration by Rational Agents, *American Political Science Review* 92: 413–418.
Merrill, S. and Grofman, B. (1999) *A Unified Theory of Voting*, Cambridge University Press.
Merrill, S., Grofman, B. and Feld, S. (1999) Nash Equilibrium Strategies in Directional Models of Two-Candidate Spatial Competition, *Public Choice* 98: 369–383.
Miller, G. and Schofield, N. (2003) Activists and Partisan Realignment in the U.S., *American Political Science Review* 97: 245–260.
Poole, K. and Rosenthal, H. (1984) U.S. Presidential Elections 1968-1980: A Spatial Analysis, *American Journal of Political Science* 28: 283–312.
Quinn, K., Martin, A. and Whitford, A. (1999) Voter Choice in Multiparty Democracies, *American Journal of Political Science* 43: 1231–1247.
Quinn, K. and Martin, A. (2002) An Integrated Computational Model of Multiparty Electoral Competition, *Statistical Science* 17: 405–419.
Rabinowitz, G, Macdonald, S. and Listhaug, O. (1991) New Players in an Old Game: Party strategy in Multiparty systems, *Comparative Political Studies* 24: 147–185.
Rakove, J. (ed.) (1999) *James Madison: Writings*, Library of America.
Schofield, N. (2001) Generic Existence of Local Political Equilibra, in M. Lasonde (ed.) *Approximation, Optimization and Mathematical Economics*, Springer, 297–308.
Schofield, N. (2003) *Mathematical Methods in Economics and Social Choice*, Springer.
Schofield, N. (2006*a*) Equilibria in the Spatial Stochastic Model of Voting with Party Activists, *The Review of Economic Design* 10: 183–203.
Schofield, N. (2006*b*) *Architects of Political Change: Constitutional Quandaries and Social Choice Theory*, Cambridge University Press.
Schofield, N. (2007*a*) The Mean Voter Theorem: Necessary and Sufficient Conditions for Convergent Equilibrium, *Review of Economic Studies* 74: 965–980.
Schofield, N. (2007*b*) Political Equilibrium with Electoral Uncertainty, *Social Choice and Welfare* 28: 461–490.
Schofield, N. et al. (1998) Multiparty Electoral Competition in the Netherlands and Germany: A Model based on Multinomial Probit, *Public Choice* 97: 257.293.
Schofield, N. and Sened, I. (2005*a*) Multiparty Competition in Israel: 1988–1996, *British Journal of Political Science* 36: 635–663.
Schofield, N. and Sened, I. (2005*b*) Modeling the Interaction of Parties, Activists and Voters: Why is the Political Center so Empty? *The European Journal of Political Research* 44:355–390.
Schofield, N. and Sened, I. (2006) *Multiparty Democracy: Elections and Legislative Politics*, Cambridge University Press.
Stokes, D. (1963) Spatial Models and Party Competition, *American Political Science Review* 57: 368–377.
Stokes, D. (1992) Valence Politics, in D. Kavanagh (ed.) *Electoral Politics*, Oxford University Press.
Strom, K. (1990) A Behavioral Theory of Competitive Political Parties, *American Journal of Political Science* 34 (3): 565–598.
Train, K. (2003) *Discrete Choice Methods for Simulation*, Cambridge University Press.