Theories of Social Choice and Elections*

Norman Schofield
Center in Political Economy,
Washington University in St.Louis, 1 Brookings Drive,
St. Louis, MO 63130
email: schofield@economics.wustl.edu

July 7, 2005

1 Social Choice Theory

The electoral models based on the early work of Hotelling (1929) and Downs (1957) essentially suppose that the motivation of parties is to win a majority of the votes or seats. A very considerable literature developed in the period up to 1973 (ably summarized by Riker and Ordeshook, 1973), focusing on two party competition and the existence of convergent equilibrium at the electoral median. McKelvey’s thesis (1972) and his first technical paper (McKelvey, 1975) made a significant contribution to this literature. A feature of this literature was that symmetry in the electoral distribution was the sufficient condition for existence. Plott’s (1967) previous analysis indicated that symmetry was also a necessary condition. Indeed, Jerry Kramer’s (1972, 1973) papers suggested that equilibria might generally not even exist. The papers by McKelvey (1976) and Schofield (1977) though independently arriving at somewhat similar conclusions on the existence of voting cycles, used entirely different formal methods. McKelvey supposed that the electoral distribution was not symmetric, so that the electoral equilibrium, or core, was empty, and then showed that disconnected preference cycles could wander throughout the preference space. (Schofield (1977) first generalized Kramer’s result by showing that there was a local electoral condition sufficient to generate voting cycles near the point, and then demonstrated

*This paper is based on research supported by NSF Grant SES 024173.
that this condition could be expected to hold somewhere, whenever two di-
mensions were involved. Extensions of these two papers (Schofield, 1978;
McKelvey 1979) showed that very stringent symmetry conditions on voter
preferences were necessary to avoid the kind of generic instability that be-
came known as chaos. Essentially, these results suggested that, with major-
ity rule, chaos in some form could be expected in two dimensions (when the
size of the society was odd) or in three dimensions, when the size was even
(Schofield, 1983). While these may be seen as the end of the Hotelling-
Downs research program, they also heralded a long line of formal explo-
rations of voting models to determine what exactly the implications were.

Perhaps the most important application of the McKelvey-Schofield (1987)
symmetry conditions for existence of a core is not to majority rule elections,
but to existence of an equilibrium in coalition bargaining, where a small
number of parties have policy preferences, and differing political weights.
As an illustration, consider the situation in the Israel Knesset after the two
elections of 1992 and 1996.[Insert Figure 1 here].

Figure 1 shows the estimated positions of thirteen parties in the Knesset
in 1996, in a two-dimensional policy space. The background to this figure
is the estimated electoral distribution of voter ideal points, obtained from
factor analysis of survey data (Arian and Shamir, 1999). The party positions
were estimated from the party manifestos, using the same factor model de-

erived from the electoral survey (Schofield and Sened, 2005a,b). Suppose
that these party positions are indeed the preferred positions of the various
party leaders, and that these leaders have "Euclidean" preferences, derived
from monotonically decreasing utility from the preferred positions. There
are three obvious post election winning coalitions: {Likud, Shas, Third Way,
NRP} with 62 seats, {Likud, Labor} with 64 seats, and {Labor, Meretz, Third
Way, Shas} with 66 seats. A coalition of {Labor, Meretz, Shas} with the sup-
port of the small Arab parties is improbable because of the unwillingness
of the Arab parties to join in coalition. Let us designate the set of winning
coalitions after this election by the symbol \( D_0 \). Given Euclidean prefer-
ences of the party leaders, each winning coalition can be associated with a
compromise set, defined as the convex hull of the preferred positions of the
leaders of the parties belonging to the coalition. However, these compromise
sets do not intersect. In terms of the spatial model there is no "core point"
(Schofield, 1986; Schofield, Grofman and Feld, 1988; Laver and Schofield,
1990). This can be shown formally by verifying that the McKelvey-Schofield
symmetry conditions are nowhere satisfied (McKelvey and Schofield, 1987).
The arguments of McKelvey (1976,1979) and Schofield (1977,1978) suggest
that in the absence of a core point, there can be cycling among possible
coalition outcomes. However, while there was indeed a degree of coalition
instability after 1996, it is more in accord with the actual events to postulate
that the outcome of political bargaining would be a lottery across a number
of different coalition governments and policy positions within the convex
hull of all the preferred positions of party leaders. Various theories of post
election bargaining have been constructed, some based on cooperative game
theory (McKelvey, Ordehook and Winer, 1978; Ordeshook and McKelvey,
1978), social choice theory (McKelvey, 1986; Schofield, 1999; Banks, Duggan
and Le Breton, 2002) or Bayesian non-cooperative game theory (Banks and
Duggan, 2000). These theories all postulate that the lottery will depend on
the party positions and on the coalition structure $D_0$ but not on the par-
cular seat shares of the parties. Note that, although Labor was the largest
party in 1996, it was unable to form a government. In fact, Netanyahu (of
Likud) won a separate prime-ministerial election against Peres (of Labor)
and formed a coalition government with Shas. The point to note about this
election is that it was the position of Shas in Figure 1 that made it pivotal
between the two possible coalitions led either by Likud or Labor.

The election of 1992 brought about a very different coalition structure,$D_1$, say. For the 1992 seat distribution, the Likud-led coalition, including
Shas, controlled only 59 seats. The estimated positions of party leader posi-
tions for 1992 imply that the convex compromise sets of all winning coalitions
intersect in the position of the Labor party leader. Because the McKelvey
Schofield symmetry conditions are satisfied at the the Labor position, this
party is at the policy core. As suggested by bargaining theory (Laver and
Schofield, 1990; Schofield, 1993, 1995; Banks and Duggan, 2000), the Labor
party, under, Rabin was able to form a minority government and implement
its declared policy position (Sened 1996; Nachmias and Sened, 1997). Thus,
under the coalition structure $D_1$, at the vector of party positions holding in
1992, the outcome of coalition negotiation was, in fact, an essentially unique
policy outcome, namely the position $z_{labor}$ of the Labor party.

Instead of assuming that parties can correctly anticipate the electoral
consequences of the choice of policy positions, we can instead assume that,
prior to the election, they estimate the probability functions

$$\{\pi_t : t = 0, \ldots, T\}$$

the family of all the possible coalition structures. Here $\pi_t(z)$ is the proba-
bility associated with $D_t$ at the vector of party positions $z$. The vote share
functions can be regarded as stochastic variables, thus exhibiting variance
about their expectations. A model of this kind can, in principle, be used
to suggest why some parties choose a leader simply to maximize vote share, while other parties focus on the effect the leader position will have on coalition bargaining.

For example, just prior to the 1996 election, it is plausible that the probabilities $\pi_0, \pi_1$ associated with $D_0, D_1$, were very close. Under $D_1$, Shas could expect a minority government under Peres, from which it would gain no government perquisites. If the electoral outcome were $D_0$, then coalition governments under either Peres or Netanyahu would both be possible. Indeed the separate prime ministerial election race was closely run, and in expectation, both of these coalitions would be equally likely. Both governments would require a majority in the Knesset for which Shas would be crucial. Obviously, coalition possibilities for Shas under $D_0$ were much more attractive than under $D_1$, so an optimal strategy for Shas would be to position itself in order to maximize $\pi_0$, conditional on the positions of the other parties. The same conclusion obviously holds for Likud. One way for Likud to maximize $\pi_0$ is to position itself so as to maximize its expected vote share vis a vis Labor. Conversely, the coalition structure $D_1$ is much more attractive for Labor, and a proxy for this is for Labor to attempt to maximize its vote share. Notice that the two situations that occurred in 1992 and 1996 were qualitatively very different. After 1996, a sequence of minimal winning or surplus coalitions formed (Riker, 1962; Laver and Schofield, 1990), whereas in 1992 a non-winning minority coalition of Labor, with the support of Meretz, was able to govern.

While the various post election bargaining theories give some insights into multiparty negotiations, they start from the given positions and strengths of the parties. To model pre-election party positioning it is necessary to construct a theory of elections that provides a plausible "non-Downsian" account of the divergent configuration of party positions illustrated in Figure 1. This is the motivation for the next section.

2 Modelling Elections.

Partly as a result of the theoretical difficulties over the non-existence of equilibria in the two party electoral model, and also because of the need to develop empirical models of voter perception and choice that involve a "stochastic" element (Aldrich and McKelvey, 1977), attention has focused on "probabilistic" vote models. A formal basis for such models is provided by the notion of "Quantal response equilibria" (McKelvey and Palfrey, 1995, 1996, 1998). In such models, behavior of each voter is modelled by a vector of
choice probabilities (Hinich 1977; Enelow and Hinich 1984; Coughlin 1992; Lin, Enelow and Dorussen, 1999; Banks and Duggan, 2005). A standard result in this class of models is that all parties converge to the electoral origin when the parties are motivated to maximize vote share (McKelvey and Patty, 2004).

However, this convergence result need not hold if there is an asymmetry in the electoral perception of the "quality" or "valence" of party leaders (Stokes, 1992). The early empirical model of Poole and Rosenthal (1984) on US Presidential elections included these valence terms and noted that there was no evidence of candidate convergence. Formal models of elections incorporating valence have been developed recently (Ansolabehere and Snyder, 2000; Groseclose, 2001; Aragones and Palfrey, 2004, 2005), but results to date have been obtained only for the two party case. This section will present a "classification theorem" for the formal probabilistic model of voter choice, in a policy space of dimension \( w \), with an arbitrary number of parties, in which party leaders exhibit differing valences. A "convergence coefficient", incorporating all the parameters of the model will be defined. It is shown that there are necessary and sufficient conditions for the existence of a pure strategy vote maximizing equilibrium at the mean of the voter distribution. When the necessary condition fails, then parties, in equilibrium, will adopt divergent positions. In general, parties whose leaders have the lowest valence will take up positions furthest from the electoral mean.

The empirical studies of voter behavior for Israel in 1992-1996, discussed in the previous section, can then be used to show that the necessary condition for party convergence fails in these elections. The equilibrium positions as obtained from the formal result, under vote maximization, are in general, comparable with, but not identical to, the estimated positions. It is suggested that the observed discrepancy can be accounted for by a more refined model that involves strategic calculations by parties with respect to post-election coalition possibilities.

The electoral model to be presented is an extension of the multiparty stochastic model of Lin, Enelow and Dorussen (1999), constructed by acknowledging the empirical asymmetries in terms of valence. The basis for this extension is the extensive empirical evidence that valence is a significant component of the judgements made by voters of party leaders. There are a number of possible choices for the appropriate game form for multiparty competition. The simplest one, which is used here, is that the utility function for party \( j \) is proportional to its vote share, \( V_j \). With this assumption, we can examine the conditions on the parameters of the stochastic model which are necessary for the existence of a pure strategy Nash equilibrium.
The key idea underlying the formal model is that party leaders attempt to estimate the electoral effects of party declarations, or manifestos, and choose their own positions as best responses to other party declarations, in order to maximize their own vote share. The stochastic model essentially assumes that party leaders cannot predict vote response precisely, but can compute the expected vote share function. In the model with “exogenous” valence, the stochastic element is associated with the weight given by each voter, \( i \), to the average perceived quality or valence of the party leader.

**Definition 1. The Stochastic Vote Model.**

The data of the spatial model is a distribution, \( \{ x_i \in X \}_{i \in N} \), of voter ideal points for the members of the electorate, \( N \), of size \( n \). As usual we assume that \( X \) is a compact convex subset of Euclidean space, \( \mathbb{R}^w \), with \( w \) finite. Each of the parties, or agents, in the set \( P = \{ 1, \ldots, j, \ldots, p \} \) chooses a policy, \( z_j \in X \), to declare. Let \( z = (z_1, \ldots, z_p) \in X^p \) be a typical vector of agent policy positions. Given \( z \), each voter, \( i \), is described by a vector \( u_i(x_i, z) = (u_{i1}(x_i, z_1), \ldots, u_{ip}(x_i, z_p)) \), where

\[
    u_{ij}(x_i, z_j) = \lambda_j - \beta \| x_i - z_j \|^2 + \epsilon_j = u_{ij}^*(x_i, z_j) + \epsilon_j \tag{1}
\]

Here \( u_{ij}^*(x_i, z_j) \) is the observable component of utility. The term \( \lambda_j \) is the “exogenous” valence of agent \( j \), \( \beta \) is a positive constant and \( \| \cdot \| \) is the usual Euclidean norm on \( X \). The terms \( \{ \epsilon_j \} \) are the stochastic errors, whose multivariate cumulative distribution function will be denoted by \( \Psi \).

There are a number of possible distribution functions that can be used. The most common assumption in empirical analyses is that \( \Psi \) is the "extreme value Type I distribution" (sometimes called log Weibull). Empirical estimation based on this assumption is known as multinomial logit (MNL). The formal "quantal response model" introduced by McKelvey and Palfrey (1995) essentially supposes that individuals make logistic errors in estimating optimal responses, where these errors are distributed by the extreme value distribution. The electoral theorem presented here is based on this assumption. An alternative assumption is that the errors are independently and identically distributed by the normal distribution (iind), with zero expectation, each with stochastic variance \( \sigma^2 \). (Lin, Enelow and Dorussen, 1999). An even more general assumption is that the stochastic error vector \( \epsilon = (\epsilon_1, \ldots, \epsilon_p) \) is multivariate normal with general variance/covariance matrix \( \Omega \). Empirical estimation based on this assumption is known as multinomial probit (MNP). See Dow and Endersby (2004) and Quinn, Martin and Whitford (PNE).
1999) for discussion of the various empirical models. Results based on the normal distribution can be obtained, but are quite difficult to demonstrate (See Schofield, (2004, 2005a, 2005b).

We assume that the valence vector

\[ \lambda = (\lambda_1, \lambda_2, ..., \lambda_p) \] satisfies \( \lambda_p \geq \lambda_{p-1} \geq \cdots \geq \lambda_2 \geq \lambda_1 \).

Because of the stochastic assumption, the behavior of voter \( i \) is modelled by a probability function \( \rho_i \). The probability that a voter \( i \) chooses party \( j \) at the vector \( z \) is

\[ \rho_{ij}(z) = \Pr[[u_{ij}(x_i, z_j) > u_{il}(x_i, z_l)], \text{ for all } l \neq j] \] (2)

\[ = \Pr[\epsilon_i - \epsilon_j < u_{ij}(x_i, z_j) - u_{il}(x_i, z_l)], \text{ for all } l \neq j \] (3)

Here \( \Pr \) stands for the probability operator generated by the distribution assumption on \( \epsilon \). The expected vote share of agent \( j \) is

\[ V_j(z) = \frac{1}{n} \sum_{i \in N} \rho_{ij}(z) \] (4)

We shall use the notation \( V : X^p \to \mathbb{R}^p \) and call \( V \) the party profile function. In the vote model it is assumed that each agent \( j \) chooses \( z_j \) to maximize \( V_j \), conditional on \( z_{-j} = (z_1, ..., z_{j-1}, z_{j+1}, ..., z_p) \).

Because of the differentiability of the cumulative distribution function, the individual probability functions \( \{\rho_{ij}\} \) are differentiable in the strategies \( \{z_j\} \). Thus, the vote share functions will also be differentiable. Let \( x^* = (1/n)\sum_i x_i \). Then the mean voter theorem for the stochastic model, asserts that the "joint mean vector" \( z_0^* = (x^*, ..., x^*) \) is a "pure strategy Nash equilibrium". Lin, Enelow and Dorussen (1999) used differentiability of the expected vote share functions, in the situation with zero valence, as well as "concavity" of the vote share functions, to assert this theorem.. They argued that a sufficient condition for the validity of the theorem was that error variance was "sufficiently large". Because concavity cannot in general be assured, we shall utilize a weaker equilibrium concept, that of "Local Strict Nash Equilibrium"(LSNE). A strategy vector \( z^* \) is a LSNE if, for each \( j \), \( z_j^* \) is a critical point of the vote function \( V_j(z_{-j}, z_j^* \cdots z_{j-1}^*, z_{j+1}, ..., z_p^*) \) and the eigenvalues of the Hessian of this function (with respect to \( z_j \)),are negative. Definition 2 gives the various equilibrium concepts used in this chapter.

**Definition 2. Equilibrium Concepts.**

(i) A strategy vector \( z^*=(z_1^*, ..., z_{j-1}^*, z_j^*, z_{j+1}^*, ..., z_p^*) \in X^p \) is a local strict pure strategy Nash equilibrium(LSNE) for the profile function \( V : X^p \rightarrow \... \)
\( \mathbb{R}^p \) iff, for each agent \( j \in P \), there exists a neighborhood \( X_j \) of \( z_j^* \) in \( X \) such that \( z_j^* \) is the strictly best response in the neighborhood \( X_j \) by \( j \) to 
\[
z^*_{-j} = (z_1^*, \ldots, z_{j-1}^*, z_{j+1}^*, \ldots, z_p^*).
\]
That is:
\[
V_j(z_1^*, \ldots, z_{j-1}^*, z_j^*, z_{j+1}^* \ldots, z_p^*) > V_j(z_1^*, \ldots, z_{j-1}^*, z_j, z_{j+1}^* \ldots, z_p^*) \quad \text{for all } z_j \in X_j - \{z_j^*\}
\]

Say the strategy \( z_j^* \) is a "local strict best response" to \( z_{-j}^* \).

We can also define local weak best response, global strict best response and global weak best response to \( z_{-j}^* \), by weakening the inequality sign, and by requiring that the response is best not just in a neighborhood, but in \( X \) itself. This allows us to define the notions of local weak pure strategy Nash equilibrium (LNE), global weak pure strategy Nash equilibrium (PNE) and global strict pure strategy Nash equilibrium (PSNE).

Obviously if \( z^* \) is an LSNE or a PNE then it must be an LNE, while if it is a PSNE then it must be an LSNE. We use the notion of LSNE to avoid problems with the degenerate situation when there is a zero eigenvalue to the Hessian. The weaker requirement of LNE allows us to obtain a necessary condition for \( z_0^* = (x^*, \ldots, x^*) \) to be a LNE and thus a PNE, without having to invoke concavity. The result below also gives a sufficient condition for the joint mean vector \( z_0^* \) 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 \( z_0^* \) 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 \( w \). Similarly, for \( z_0^* \) to be a LNE, then the convergence coefficient must be weakly bounded above by \( w \). The main point of the result is that when this condition fails, then the joint mean vector \( z_0^* \) cannot be a LNE and therefore cannot be a PNE. Of course, even if the sufficient condition is satisfied, and \( z_0^* = (x^*, \ldots, x^*) \) is an LSNE, it need not be a PNE.

To state the theorem, we first transform coordinates so that in the new coordinates, \( x^* = 0 \). We shall refer to \( z_0 = (0, 0) \) 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 covariation matrix. We first define this, and then show that the vote share Hessians depend on the covariation matrix.

**Definition 3:** The electoral covariance matrix, \( \frac{1}{n} \nabla \). To characterize the variation in voter preferences, we represent in a simple form the covariation matrix (or data matrix), \( \nabla \), given by the distribution of voter ideal
points. Let \( X \) have dimension \( w \) and be endowed with a system of coordinate axes \((1, \ldots, r, s, \ldots, w)\). For each coordinate axis let \( \xi_r = (x_{1r}, x_{2r}, \ldots, x_{nr}) \) be the \( n \)-vector of the \( r^{th} \) coordinates of the set of \( n \) voter ideal points. We use \((\xi_r, \xi_s)\) to denote scalar product.

The symmetric \( w \times w \) voter covariation matrix \( \nabla \) is then defined to be the matrix \( \nabla = [(\xi_r, \xi_s)] \), while the covariance matrix is defined to be \( \frac{1}{n} \nabla \).

We write \( \nu^2_r = \frac{1}{n}(\xi_r, \xi_r) \) for the electoral variance on the \( r^{th} \) axis and

\[
v^2 = \sum_{r=1}^{w} \nu^2_r = \frac{1}{n} \sum_{r=1}^{w} (\xi_r, \xi_r) = \text{trace}(\frac{1}{n} \nabla)
\]

for the total electoral variance. The electoral covariance between the \( r^{th} \) and \( s^{th} \) axes is \((\nu_r, \nu_s) = \frac{1}{n}(\xi_r, \xi_s)\).

**Definition 4: The Extreme Value Distribution, \( \Psi \).**

(i) The cumulative distribution \( \Psi \) and probability density function \( \varphi \) have the closed forms

\[
\Psi(h) = \exp[-\exp[-h]], \\
\varphi(h) = \exp[-h] \exp[-\exp[-h]],
\]

with variance \( \frac{1}{6} \pi^2 \).

(ii) With this distribution it follows from Definition 2 that that, for each voter \( i \), and party \( j \), the probability \( \rho_{ij}(z) \) is given by the logistic quantal response expression

\[
\rho_{ij}(z) = \frac{\exp[u^*_ij(x_i, z_j)]}{\sum_{k=1}^{p} \exp[u^*_ik(x_i, z_k)]}, \quad (5)
\]

Note that (ii) implies that the model satisfies the independence of irrelevant alternative property (IIA): for each individual \( i \), and each pair \( j, k \), the ratio \( \rho_{ij}(z)/\rho_{ik}(z) \) is independent of a third party \( l \) (See Train, 2003,p.79)

While this distribution assumption facilitates estimation, the IIA property may be violated. It is possible to obtain results for the case of covariant errors, so that IIA is not imposed (Schofield, 2005a).

The formal model just presented, and based on \( \Psi \) is denoted \( M(\lambda, \beta; \Psi, \nabla) \), though we shall usually suppress the reference to \( \nabla \).
It can then easily be shown that, at the vector $\mathbf{z}_0 = (0, \ldots, 0)$, the probability $\rho_{ij}(\mathbf{z}_0)$ that individual $i$ votes for party $j$ is the same for every individual, and is given by

$$\rho_j = \left(1 + \sum_{k \neq j} \exp[\lambda_k - \lambda_j]\right)^{-1}$$

Then the Hessian of the vote share function of the lowest valence party, 1, at $\mathbf{z}_0$ is given by the symmetric matrix

$$C_1 = \left[2[1 - 2\rho_1]\left(\frac{\beta}{n} \nabla\right) - I\right]$$

where $I$ is the $w$ by $w$ identity matrix. Since it is trivial to show that the first order conditions, $\frac{\partial V_j}{\partial z_j} = 0$ are satisfied at $\mathbf{z}_0$, it follows that the necessary condition for existence of a LSNE at $\mathbf{z}_0$ is that $C_1$ has negative eigenvalues. Moreover, if $C_1$ has negative eigenvalues at $\mathbf{z}_0$, then so will the Hessians for $j = 2, \ldots, p$. Thus we obtain necessary and sufficient conditions in terms of $C_1$. Because this condition is determined by the determinant and trace of $C_1$, it can be re-expressed in terms of a convergence coefficient

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

Then a sufficient condition for existence of a LSNE at $\mathbf{z}_0$ is that

$$c(\lambda, \beta; \Psi, \nabla) < 1.$$  

Moreover, in dimension $w$, the necessary condition for existence of a LNE at $\mathbf{z}_0$ is $c(\lambda, \beta; \Psi, \nabla) \leq w$.

The definition of $\rho_1$ follows directly from the extreme value distribution. Obviously if all valences are identical then $\rho_1 = \frac{1}{w}$, as expected. The effect of increasing $\lambda_j$, for $j \neq 1$, or of decreasing $\lambda_1$ is clearly to decrease $\rho_1$, and therefore to increase $c(\lambda, \beta; \Psi, \nabla)$, thus rendering existence of a LSNE less likely. Ceteris paribus, a LNE at the joint origin is “less likely” the greater are the parameters $\beta$, $\lambda_p - \lambda_1$ and $\nu^2$. The proof of this result is given in Schofield (2005a). Note that for a general spatial model with an arbitrary, non-quadratic but differentiable metric $\Pi(x_i, z_j) = ||x_i - z_j||$, a similar expression for $c(\lambda, \beta; \Psi, \nabla)$ can be obtained, but in this case the covariance matrix $\frac{\beta}{n} \nabla$ will depend on $\Pi$. Note also that if the non-differentiable Cartesian metric $\Pi(x_i, z_j) = \sum_{k=1}^w |x_{ik} - z_{jk}|$ were used, then the first order condition would be satisfied at the median rather than the mean.

10
Even when the sufficient condition is satisfied, so the joint origin is an LSNE, the concavity condition (equivalent to the negative semi definiteness of all Hessians everywhere) is so strong that there is no good reason to expect it to hold. The empirical analyses of Israel, which we shall present below, show that the necessary condition fails. In this polity, a vote maximizing PNE, even if it exists, will generally not occur at the origin. In these analyses, the policy space is two-dimensional, and in this case it is possible to demonstrate that the eigenvalues, $a_1, a_2$ of the Hessian of the lowest valence party, 1, are given by the expressions

$$a_{1,2} = [1 - 2\rho_1]/\beta\{[v_1^2 + v_2^2] ± [v_1^2 - v_2^2]^2 + 4(v_1, v_2)^2\}^{\frac{1}{2}} - 1$$

Note that the case $\lambda_p = \lambda_1$ was studied by Lin, Enelow and Dorussen (1999), under the assumption that the errors were independently and identically normally distributed. A similar result to the above can be obtained for this formal model based on multivariate normal errors (Schofield, 2004). The only difference is that with the normal distribution the convergence coefficient has the error variance $\sigma^2$ in the denominator, and has the average of the valence difference in the numerator. It follows that if all valences are identical, then the average valence difference is zero, and thus the electoral origin is assured of being a LSNE. However, this does not guarantee that it is a PSE. However, if the error variance is sufficiently great (in comparison to the spatial coefficient, $\beta$, and the electoral variance, $v^2$) then all Hessians will be negative definite everywhere. This implies that the joint origin will indeed be a PSE.

In the next section we use this result to determine whether convergence can be expected in the complex multiparty situation in the Israel Knesset.

### 3 Empirical Analysis for Israel: the election of 1996.

To provide an explanation for the non-convergent positions of the parties at the time of the 1996 election, a MNL estimation of the election based on the Arain-Shamir survey was carried out. Details are given in Schofield and Sened (2005b) and Schofield, Sened and Nixon, (1998). The MNL model with valence was found to be statistically superior to both a MNL model and a multinomial probit (MNP) model without valence. The two dimensions of policy deal with attitudes to the PLO (the horizontal axis) and religion (the vertical axis). The policy space was derived from voter surveys (obtained by Arian and Shamir, 1999) and the party positions from analysis of party
manifestos. Using the formal analysis, we can readily show that one of the eigenvalues of the lowest 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 Figure 2.

In 1996, the lowest valence party was the NRP with valence –4.52, while the valences for the major parties, Labor and Likud are 4.15 and 3.14 respectively. The spatial coefficient was \( \beta = 1.12 \), so for the extreme value model \( M(\Psi) \) we compute \( \rho_{NRP} \approx 0 \). Since \( v_1^2 = 1.0, v_2^2 = 0.732 \) and \( (v_1, v_2) = 0.591 \), we can compute \( C_{NRP} \) as follows:

\[
\rho_{NRP} \approx \frac{1}{1 + e^{4.15+4.52} + e^{3.14+4.52}} \approx 0.
\]

\[
C_{NRP} = 2(1.12) \begin{pmatrix} 1.0 & 0.591 \\ 0.591 & 0.732 \end{pmatrix} - I = \begin{pmatrix} 1.24 & 1.32 \\ 1.32 & 0.64 \end{pmatrix}
\]

\[
c(\lambda, \beta; \Psi, \nabla) = 3.88
\]

Using the expression for the eigenvalues presented above, we find that the eigenvalues are 2.28 and -0.40, giving a saddlepoint, and a value for the convergence coefficient of 3.88, well in excess of the necessary bound of 2.0. 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,-1.25) and on this axis the NRP vote share decreases. Figure 2, gives one of the local equilibria in 1996, obtained by simulation of the model. The Figure makes it clear that, as predicted by the formal model, all vote maximizing positions lay on the principal axis through the origin and the point (1.0,0.8). Five different LSNE were located. However, in all the equilibria, the two high valence parties, Labor and Likud, were located at precisely the same positions. The only difference between the various equilibria were that the positions of the low valence parties were perturbations of each other. [Insert Figure 2 about here]

Analyses and simulations for 1992 and 1988 gave figures very similar to that of Figure 2.

The simulated vote maximizing party positions in all three elections indicated that there was no deviation by parties off the principal 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. In particular, the simulation confirms the logic of the above 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 from the origin.

In contrast to these low valence parties, because the valence differences between Labor and Likud were relatively low in all three elections, their equilibrium 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. 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 $r^2 \approx 0.47$). 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 principal eigenvector, (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 to 1996. Indeed, for 1996, Figure 1 indicates that there was evidence of bifurcation in the electoral distribution.

In comparing Figure 1, of the estimated party positions, and Figure 2, of simulated equilibrium positions, there is a notable disparity particularly in the position of Shas between the estimated party positions of Figure 1, and the simulated equilibrium positions in Figure 2. As we have observed, in 1996, Shas was pivotal between Labor and Likud, in the sense that to form a winning coalition government, both of the two larger parties required the support of Shas. It is obvious that the location of Shas in Figure 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 adopted a strategic position to make use of his pivot
advantage.

The relationship between the empirical work and the formal model, together with the possibility of strategic reasoning of this kind, suggests the following conclusion.

**Conclusion.** The close correspondence between the simulated LSNE based on the empirical analysis and the estimated actual political configuration suggests that the true utility function for each party \( j \) has the form

\[
U_j(z) = V_j(z) + \delta_j(z)
\]

where \( \delta_j(z) \) may depend on the beliefs of party leaders about the post election coalition possibilities, as well as the effect of activist support for the party. This conclusion leads to the further conjecture, for the set of feasible strategy profiles in the Israel polity, that \( \delta_j(z) \) is “small” relative to \( V_j(z) \). A formal model to this effect could indicate that the LSNE for \( \{U_j\} \) would be close to the LSNE for \( \{V_j\} \).

The discussion of coalition bargaining offered in the first section of this chapter suggests that the primary motivation of high valence parties can be assumed to be to maximize the probability of the bringing about the more favorable coalition structure, and a proxy for this can be taken to be expected vote share. Low valence parties like Shas or NRP may have more complex motivations, involving positioning themselves to be able to bargain effectively over coalition formation. Thus, to construct a formal model of political behavior it would appear necessary to combine elements of social choice, as discussed in the first section, together with a theory of electoral competition.

The construction of a formal model involving both electoral and coalitional concerns would have appealed to Richard McKelvey. Such a theory requires integrating social choice theory, multinomial logit electoral models based on quantal response and factor analysis, together with computation and simulation of the equilibria (McKelvey and McLennan, 1996).

### 4 General Conclusions: Party Switching

In an attempt to model this complex political game, this chapter has introduced the idea of a *local Nash equilibrium* (Schofield and Sened, 2002, 2006 a,b,c). This general concept can incorporate the quantal response idea for modelling electoral response, as developed by McKelvey and Palfrey (1995) and McKelvey and Patty (2004) as well as the spatial notions of equilibrium studied in the earlier analyses of McKelvey (1976,1979) and Schofield (1977,1978) . The underlying premise of this concept is that party principals will not consider "global" changes in party policies, but will instead propose
small changes in the party leadership in response to changes in beliefs about
electoral response and the likely consequences of policy negotiations. It is
also evident that the electoral model depends on the notion of "valence".
Valence can be regarded as that element of a voter’s choice which is de-
termined by judgement rather than preference. This accords well with the
arguments of James Madison in "Federalist 10" of 1787 and of Condorcet
in his treatise of 1785 on social choice theory (Schofield, 2005c), and can be
interpreted as a justification for the democratic process. Pursuing this line
of development suggests a new and very interesting way of thinking about
politics in terms of "belief games". Indeed, it is natural to speculate that
political processes, regarded as belief games, are in some kind of dynamic
equilibrium (Miller and Schofield, 2003) between the opposed convergent
and centripetal tendencies implicit in the models that we have discussed
here.

The general conclusions about party switching and pre-election agree-
ments that can be drawn from the formal models are:

(i) Under proportional rule, and with exogeneous valence alone, there will
be a centrifugal tendency associated with low valence parties which may
make them disagreeable partners to high valence parties. This follows
because the addition of a low valence party will surely lower the valence of
the higher valence party. If activists are relevant, then their influence
would be to pull the party away from the center.

(ii) Two low valence parties may combine, but this will have little effect
on their combined valence, and their optimal position will still be radical.

The coalition effect may be pronounced however. Neither party may be
pivotal, but by combining they may become, like Shas in Israel, influential
in coalition formation.

If they cater to different activist groups, then there will be conflict be-
tween the groups over choice of leader. The analyses for the Conservative
Party in Britain suggested there was conflict between anti-Europe activists
and pro Capital activists.

The evidence from the Netherlands indicates that activists do influence
party position. According to Theorem 4, this effect should be most pro-
nounced for parties whose leaders have relatively low exogeneous valence.

(iii) Coalescence between high valence parties may occur. In Israel,
Sharon appears willing to form an electoral pact with Labor. This change
in the distribution of electoral power could induce the new party to adopt a
centrirst position.

(iv) Because low valence parties may resent the dominance of a core
party, they may adjust position, or recreate themselves in order to destroy the center party dominance. This appears to have occurred in 1992 in Italy (Mershon, 2002; Schofield, 1993, Giannetti and Sened 2004).

These are only a few of the inferences that may be drawn on the basis of the various models presented here.

5 References


19


Figure 1: Party Positions in the Israel Knesset 1996.
Figure 2: Equilibrium Party Positions in the Knesset
Key: 1=Shas, 2=Likud, 3=Labor, 4=NRP, 5=Molodet, 6=Meretz and Third Way