John Deegan

1976

A MULTI-DIMENSIONAL STATISTICAL PROCEDURE FOR SPATIAL ANALYSIS*

Lawrence Cahoon
Virginia Polytechnic Institute

Melvin J. Hinich†
Virginia Polytechnic Institute

Peter C. Ordeshook††
Carnegie-Mellon University

†Present address: Fairchild Distinguished Scholar, California Institute of Technology

††Center for Advanced Study in the Behavioral Sciences, Stanford

A MULTI-DIMENSIONAL STATISTICAL PROCEDURE FOR SPATIAL ANALYSIS*

The methodological theory and application discussed in this paper is motivated by the spatial theory of electoral competition. Specifically, we develop a statistical procedure that recovers from cross-sectional survey thermometer score data the spatial configuration of candidates and citizens, as well as the dimensionality of the issue space and the relative salience of these issues.

Briefly, spatial theory seeks to ascertain the policies candidates adopt in an election. The principal assumption of the theory is that candidate strategies and citizen preferences can be represented in an n-dimensional Euclidean coordinate system with a Euclidean distance metric used to describe citizen utility functions.[1] The dimensions are interpreted, then, as issues and a candidate's strategy is a position on each of the issues. We cannot review here all of spatial theory's conclusions. Its best known result, however, is that, under a variety of assumptions about citizen preferences and candidate objectives, candidates should adopt policies at or near the electorate's median preference on each dimension.[2]

Empirical research that addresses this theory directly or indirectly, however, focuses on its assumptions rather than on its deductions. That is, this research is concerned with testing the empirical adequacy of alternative assumptions about nonvoting, the

relevance of "issues" for citizen choice, and the degree to which choices appear "rational" because of cognitive balance.[3] This gap between theory and empirical research can be attributed to the apparent necessity for a multi-dimensional accounting of preference and strategies and, in turn, to the severe methodological difficulties associated with ascertaining dimensions of preference and the citizens' perceptions of candidates on them.

Stemming from a vast literature on multi-dimensional scaling, political scientists are beginning to address this problem -- the most relevant examples being Rusk and Weisberg's analysis of SRC thermometer data, Daalder and Rusk's analysis of Dutch elite data, and Rabinowitz's nonmetric algorithm for recovering candidate spatial positions from mass survey data.[4] With the exception of Rabinowitz, however, the scaling literature pays little if any attention to statistical matters.[5] Thus, our principal objective here is not simply to develop an alternative algebraic scaling model but rather to formulate a statistical methodology for estimating the parameters of a theoretical model of election competition.

We present that methodology in the next section. Briefly, taking the data analyzed by Rusk and Weisberg and by Rabinowitz, we assume that thermometer scores are derived from a weighted Euclidean distance metric. Since it is known that, with this specification, a straightforward factor analysis of the data is inappropriate, we modify the data in a way that closely parallels the techniques developed by Torgerson and by Schoneman.[6] Since noisy data are assumed explicitly in our analysis, however, an adequate treatment of that noise is one of

our principal concerns (a complete discussion of the method necessarily entails considerable mathematical complexities that obscure its fundamental structure and, hence, we periodically limit our exposition to special cases while relegating the proof of two results to an appendix). In the second section we examine the effects of valence issues and of candidates who are perceived as extremists by a significant proportion of citizens. In the third section we test the robustness of the methodology and its sensitivity to the assumptions with artificial data. We conclude with an application to the thermometer data collected in the University of Michigan Survey Research Center's 1968 election survey.

A Statistical Model

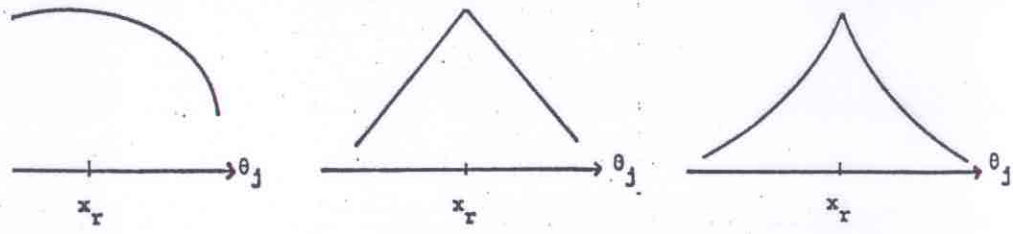
Generally, spatial theory begins with the assumption that there exists a cardinally defined issue space common to all respondents and that individual preferences can be represented by the weighted Euclidean distance metric:

$$\| \underset{\sim}{\theta}_j - \underset{\sim}{x}_r \|_A^{2/k} = [(\underset{\sim}{\theta}_j - \underset{\sim}{x}_r)' \underset{\sim}{A} (\underset{\sim}{\theta}_j - \underset{\sim}{x}_r)]^{1/k}$$

$$= [\sum_{i=1}^{n} \sum_{h=1}^{n} a_{ih}(\theta_{ji} - x_{ri})(\theta_{jh} - x_{rh})]^{1/k}$$

where $\underset{\sim}{\theta}_j = (\theta_{j1},\ldots,\theta_{jn})'$ denotes the position of candidate j on each of n dimensions, $\underset{\sim}{x}_r = (x_{r1},\ldots,x_{rn})'$ denotes the rth respondent's ideal point, and $\underset{\sim}{A} = (a_{ih})$ is a positive definite nxn matrix of issues weights. The usual assumption of spatial theory is that k = 1, i.e., if we interpret $-\| \underset{\sim}{\theta}_j - x_r \|_A^2$ as a utility function, then utility is inversely

measured by squared distance. The assumption of a common cardinal space, however, is problematic since "issues" typically possess no natural scale. Thus, respondents may define the space with a particular sensitivity to positions that are "far" from their ideal points or alternatively they may be sensitive, i.e., perceive differences as substantively meaningful, only if positions are "near" their ideal points.[7] To accomodate the several possibilities, we permit k to vary. If, for example, each respondent is more sensitive to positions that are far from $\underset{\sim}{x}_r$, then to place all respondents in a common space, we let k = 1 (see Figure 1a). If the respondents' sensitivity is uniform, k = 2 (see Figure 1b); if respondents are more cognizant of differences near their ideal points, then we let k = 4 (see Figure 1c). Since we do know the appropriate value for k in a given population, k is a parameter that we must estimate.



(a) k = 1 (b) k = 2 (c) k = 4

Figure 1: Shape of Distance Metric Used to Represent Preferences in a Common Euclidean Space

Suppose now that N respondents rate each of p + 1 candidates on a thermometer scale (that varies from 0 to 100), where

the number of candidates exceeds the number of issues ($p > n$), and that respondent r's thermometer score for candidate j takes the form

$$T_{jr} = 100 - D_{jr}$$

where

$$D_{jr} = [(\underset{\sim}{\theta}_j - \underset{\sim}{x}_r)' \underset{\sim}{A}(\underset{\sim}{\theta}_j - \underset{\sim}{x}_r)]^{1/k} + \varepsilon_{jr} \qquad (1)$$

and where ε_{jr} is a stochastic error term. Since we can subtract observed thermometer scores from 100, we work with expression (1).

It is self-evident that we should make allowance for noise and lack of fit in the model (viz ε_{jr}). In fact, Rabinowitz notes that one of the major difficulties associated with applying standard scaling procedures to thermometer data is the sensitivity of these procedures to noise and the inherent noisiness of such data.[8] Presently, we assume that the errors are distributed with zero mean, are across candidates as well as across respondents, and that the variances of the errors are equal for all candidates. While there are reasons for believing that these assumptions will not be satisfied in general, they are the usual and a convenient starting point for developing statistical methodologies. We note, moreover, that scaling techniques which fail to consider the structure of noise in their data at best implicitly require this assumption.

Assuming that the scoring experiment represented by expression (1) is performed by each respondent in a sample of the electorate, our goal is to identify (estimate) the dimensionality of the issue space, n,

each candidate's position in the space, $\underset{\sim}{\theta}_j$, each citizen's ideal point, $\underset{\sim}{x}_r$, and the matrix of issue weights, $\underset{\sim}{A}$. Our procedure is to appropriately modify the data (thermometer scores) and apply the principal components version of factor analysis. That this data must be modified in some way follows from the fact that factor analysis assumes a linear model whereas the model of expression (1) is nonlinear (owing to the term $\underset{\sim}{x}_r'\underset{\sim}{A}\underset{\sim}{x}_r$).

We begin with two assumptions: (1) all citizens have the same perception of each candidate; and (2) each a_{ih} varies independently of $\underset{\sim}{x}_r$ in the population, e.g., all citizens weight the issues in an identical fashion. Clearly, these two assumptions, while implicit in metric scaling techniques and while possessing a long history in spatial analysis, are restrictive. Without the assumption of some structure, however, estimation is impossible. As a partial resolution of the problem, then, when analyzing the SRC thermometer data we divide the sample by partisian identification to check whether there are significant differences in $\underset{\sim}{A}$ or in the $\underset{\sim}{\theta}$'s. Of course, there are many interesting partitions of the sample which could be analyzed using the methodology.

Using only thermometer data to develop a space, there is necessarily an ambiguity between $\underset{\sim}{A}$ and the covariance matrix $\underset{\sim}{\Sigma}$ of the population ideal points. That is, the units of A are defined in terms of $\underset{\sim}{\Sigma}$ and vise-versa, or, more precisely, $\underset{\sim}{A}^{1/2}\underset{\sim}{\Sigma}\underset{\sim}{A}^{1/2}$ can be estimated but not $\underset{\sim}{A}$ and $\underset{\sim}{\Sigma}$ separately. Without loss of generality, then, we can set $\underset{\sim}{\Sigma}$ equal to the identity matrix $\underset{\sim}{I}$ and estimate $\underset{\sim}{A}$. This simply defines $\underset{\sim}{A}$ in terms of standard deviation units. In our

model, however, we must impose the restricting assumption that A is diagonal (i.e., $a_{ih} = 0$ for all $i \neq h$) so that our procedure estimates a_{ii}, $i = 1, \ldots, n$ (hereafter, we simplify notation by letting $a_i = a_{ii}$). It is not surprising that constrants have to be placed on the space since there are no natural units to the axes.

We proceed, now, to construct a new set of observations that are linear in $\underset{\sim}{\theta}_j$ and $\underset{\sim}{x}_r$.[9] To this end we first raise the original observations to the k^{th} power and obtain

$$D_{jc}^{k} = \underset{\sim}{\theta}_j'\underset{\sim\sim}{A}\underset{\sim}{\theta}_j - 2\underset{\sim}{\theta}_j'\underset{\sim\sim}{A}\underset{\sim}{x}_r + \underset{\sim}{x}_r'\underset{\sim\sim}{A}\underset{\sim}{x}_r + U_{jr} \qquad (2)$$

where

$$U_{jr} = \sum_{\ell=1}^{k} \binom{k}{\ell} \| \underset{\sim}{\theta}_j - \underset{\sim}{x}_r \|^{k-\ell} \varepsilon_{jr}^{\ell} \qquad (3)$$

At this point we can simplify exposition by restricting the analysis to the case of $k = 1$ and by relegating appropriate generalizations to the appendices (in particular, Results 1 and 2 which follow). For $k = 1$ expression (2) then becomes

$$D_{jr} = \underset{\sim}{\theta}_j'\underset{\sim\sim}{A}\underset{\sim}{\theta}_j + 2\underset{\sim}{\theta}_j'\underset{\sim\sim}{A}\underset{\sim}{x}_r + \underset{\sim}{x}_r'\underset{\sim\sim}{A}\underset{\sim}{x}_r + \varepsilon_{jr} \qquad (4)$$

Without loss of generality we set $\underset{\sim}{\theta}_{p+1} = 0$ since the thermometer scores derived from expression (1) are invariant with respect to the location of the origin, i.e., we let

$$D_{p+1,r} = \underset{\sim}{x}_r'\underset{\sim\sim}{A}\underset{\sim}{x}_r + \varepsilon_{p+1,r} \qquad (5)$$

Thus by subtracting $D_{p+1,r}$ from D_{jr}, we eliminate the nonlinear term. To eliminate the term $\underset{\sim}{\theta}_j'\underset{\sim\sim}{A}\underset{\sim}{\theta}_j$, observe that

$$\bar{D}_j = \frac{1}{N}\sum_{r=1}^{N} D_{jr} = \underset{\sim}{\theta}_j'\underset{\sim\sim}{A}\underset{\sim}{\theta}_j - 2\underset{\sim}{\theta}_j'\underset{\sim\sim}{A}\underset{\sim}{\bar{x}} + \underset{\sim}{\bar{x}}'\underset{\sim\sim}{A}\underset{\sim}{\bar{x}} + \bar{\varepsilon}_{jr} \qquad (6)$$

where the overbars indicate averages over the sample of respondents.

Thus letting

$$Y_{jr} = D_{jr} - D_{p+1,r} - [\bar{D}_j - \overline{D_{p+1}}] \qquad (7)$$

the new set of observations becomes, from expressions (5), (6), and

$$Y_{jr} = -2\underset{\sim}{\theta}_j'\underset{\sim\sim}{A}(\underset{\sim}{x}_r - \underset{\sim}{\bar{x}}) + (\varepsilon_{jr} - \bar{\varepsilon}_j) - (\varepsilon_{p+1,r} - \overline{\varepsilon_{p+1}}) \qquad (8)$$

which is linear in $\underset{\sim}{\theta}_j$ and $\underset{\sim}{x}_r$, i.e., we now possess a set of observations whose covariance matrix computed across respondents can be factored.

If the model were perfect and explained the data with no error, the pxp covariance matrix of the Y_{jr} would be $4\underset{\sim}{\theta}'\underset{\sim}{A}^2\underset{\sim}{\theta}$, where $\underset{\sim}{\theta}$ is the nxp matrix whose jth column is the jth candidate's

position θ_j. Since the $n \times n$ matrix A^2 is diagonal, factoring the covariance matrix of the Y_{jr} yields a rotation of the $n \times p$ matrix $\underset{\sim}{A}\underset{\sim}{\theta}$. The positive weight a_i stretches or shrinks the positions of the candidates on the ith axis ($i = 1, \ldots, n$) depending a whether a_i is greater or less than one standard deviation of x_{ri}. We later show how to estimate these a_i weights for $n = 2$.

Not surprisingly the model fails to be perfect for many reasons. Even if our model with an additive error term was a good approximation, the covariance matrix of the Y_{jr} is not simply $4\underset{\sim}{\theta}'\underset{\sim}{A}^2\underset{\sim}{\theta}$ plus a small error. In the appendix we derive the covariance matrix of the Y_{jr} for $k = 4$ -- the relevant case for our empirical analysis -- but for sake of exposition we now state the result for $k = 1$.[10]

Result 1: Suppose that the ε_{jr} errors are uncorrelated across both respondents and candidate scores, have zero means, and common variance σ^2. If the $n \times n$ covariance matrix of the ideal points $\underset{\sim}{\Sigma}$ is set equal to the identity matrix, then the sample covariance matrix of the Y_{jr} for $k = 1$ is

$$\frac{1}{N}\underset{\sim}{Y}'\underset{\sim}{Y} = 4\underset{\sim}{\theta}'\underset{\sim}{A}^2\underset{\sim}{\theta} + \sigma^2\underset{\sim}{M} + \underset{\sim}{\varepsilon}$$

where $\underset{\sim}{M}$ is the nonsingular $p \times p$ matrix whose diagonal elements are all two and whose off-diagonal elements are all one, and $\underset{\sim}{\varepsilon}$ is a $p \times p$ matrix of errors whose magnitudes are approximately $N^{-1/2}$.

When N, the number of respondents, is large, we could obtain a good estimate of a rotation of $\underset{\sim}{A}\underset{\sim}{\theta}$ be factoring

$$\frac{1}{N}\underset{\sim}{Y}'\underset{\sim}{Y} - \sigma^2\underset{\sim}{M}$$

if we knew σ^2. We can obtain an estimate of σ^2 by using the algebraic fact that the matrix $\underset{\sim}{\theta}'\underset{\sim}{A}^2\underset{\sim}{\theta}$ has only $n < p$ nonzero eigenvalues since it has rank n. Thus the $p - n$ smallest eigenvalues of

$$\frac{1}{N}\underset{\sim}{Y}'\underset{\sim}{Y} - \sigma^2\underset{\sim}{M}$$

depends on the elements of $\underset{\sim}{\varepsilon}$, which are approximately $N^{-1/2}$ for large N.

Result 2: If σ^2 is estimated by choosing $\hat{\sigma}^2$ such that the smallest eigenvalue of the $p \times p$ matrix

$$\hat{W} = \frac{1}{N}\underset{\sim}{Y}'\underset{\sim}{Y} - \hat{\sigma}^2\underset{\sim}{M}$$

is zero, then

1) $\hat{\sigma}^2$ exists and is unique,
2) $\hat{\sigma}^2 - \sigma^2$ is approximately $N^{-1/2}$ for large N,

and

3) $\hat{W} = 4\underset{\sim}{\theta}'\underset{\sim}{A}^2\underset{\sim}{\theta} + O(N^{-1/2})$

where the elements of $O(N^{-1/2})$ are approximately $N^{-1/2}$ for large N.[11]

The matrix $\hat{W}$ / 4 is then factored by principal components to obtain an estimate of $\Gamma'A\theta$, where Γ is some unknown rigid rotation of the coordinate system, i.e., $\Gamma'\Gamma = I$. Thus we need to estimate the diagonal elements of A and the rotation angles in the Γ elements in order to estimate the candidate positions θ. In contrast to the situation in factor analysis, Γ is estimable subject to some symmetry ambiguities.[12]

In order to proceed with the parameter estimation, the dimensionality of the space must be determined. This is done by studying the eigenvalue pattern of the sample covariance matrix $\frac{1}{N}Y'Y$ and, more importantly, the eigenvalues of the adjusted matrix $\hat{W}$. The problems of selecting the dimensionality of the space and identifying the dimensions in terms of real issues is similar to that encountered using factor analysis. Even when other issue and perceptual data is available, there is no mechanical substitute for scientific judgment and experience.

The estimation of the issue weights and of the rotation makes use of the $n \times p$ matrix, denoted $\hat{F}$, which is the factor of $\hat{W}$. As $N \to \infty$ this matrix converges to the matrix $F = \Gamma A\theta$. Consequently,

$$\theta'A\theta = F'\Gamma'A^{-1}\Gamma F + O(N^{-1/2}) \qquad (9)$$

For the jth candidate, expression (9) becomes

$$\theta_j'A\theta_j = \sum_{i=1}^{n}\sum_{h=1}^{n} a_i^{-1}(\gamma_{ih}\hat{f}_{hj})^2 + O(N^{-1/2}), \qquad (10)$$

where γ_{ih} is the ihth element of Γ, $\hat{f}_{jh}$ is the jhth element of $\hat{F}$, and similarly

$$\theta_j'A\bar{x} = \sum_{i=1}^{n}\sum_{h=1}^{n} \bar{x}_i\gamma_{ih}\hat{f}_{hj} + O(N^{-1/2}). \qquad (11)$$

The sample mean ideal position on the ith issue, $\bar{x}_i$, is unobserved and must be estimated. Finally, we have from (7), (10) and (11) that

$$\overline{D_j} - \overline{D_{p+1}} = \sum_{i=1}^{n}\left[a_i^{-1}\left(\sum_{h=1}^{n}\gamma_{ih}\hat{f}_{hj}\right)^2 - 2\bar{x}_i\left(\sum_{h=1}^{n}\gamma_{ih}\hat{f}_{hj}\right)\right] + O(N^{-1/2}). \qquad (12)$$

We thus have p equations with $3n-1$ unknowns (the a_i^{-1}'s, the $\bar{x}_i$'s, and the γ_{ih}'s which depend on $n-1$ rotation angles).[13]

The method for estimating these unknowns and using them to get the θ_j position is most easily shown for the case where $n = 2$. For $n = 2$, all orthogonal rotation matricies are of the form (disregarding reflection across either or both axes),[14]

$$\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} = \begin{bmatrix} \cos\xi & -\sin\xi \\ \sin\xi & \cos\xi \end{bmatrix}$$

Substituting this form into (12) yields

$$\overline{D_j^2} - \overline{D_{p+1}^2} = \alpha_0\hat{f}_{j1}^2 + \alpha_1\hat{f}_{j2}^2 + \alpha_2\hat{f}_{j1}\hat{f}_{j2} + \alpha_3\hat{f}_{j1} + \alpha_4\hat{f}_{j2} + O(N^{-1/2}) \quad (13)$$

where

$$\alpha_0 = a_1^{-1}\cos^2\xi + a_2^{-1}\sin^2\xi \quad (14)$$

$$\alpha_1 = a_1^{-1}\sin^2\xi + a_2^{-1}\cos^2\xi \quad (15)$$

$$\alpha_2 = 2(a_2^{-1} - a_1^{-1})\sin\xi\cos\xi \quad (16)$$

$$\alpha_3 = -2(\bar{x}_1 a_1^{-1/2}\cos\xi + \bar{x}_2 a_2^{-1/2}\sin\xi) \quad (17)$$

$$\alpha_4 = -2(-\bar{x}_1 a_1^{-1/2}\sin\xi + \bar{x}_2 a_2^{-1/2}\cos\xi) \quad (18)$$

Ordinary least squares yields consistent estimates of α_0 through α_4. Using these estimates in expressions (14), (15), and (16), we solve for ξ, a_1, and a_2 as follows:[15]

$$\tan 2\xi = \frac{\alpha_2}{\alpha_1 - \alpha_0}, \quad (19)$$

$$2a_1^{-1} = \alpha_1 + \alpha_0 - (\sin 2\xi)^{-1}\alpha_2, \quad (20)$$

and

$$2a_2^{-1} = \alpha_1 + \alpha_0 + (\sin 2\xi)^{-1}\alpha_2. \quad (21)$$

From (19) we can derive directly an estimate of $\underline{\Gamma}$. Further, expressions (19), (20), and (21) provide an estimate of $\underline{A}$, while these results in conjunction with (17) and (18) permit us to calculate $\bar{x}_i$. Finally, the estimate of $\underline{\theta}$ is given by $\hat{\underline{\theta}} = \hat{\underline{A}}^{-1}\hat{\underline{\Gamma}}\,\hat{\underline{F}}$.

Once the candidate positions and the other parameters of the space are estimated, it is easy to obtain an estimate of the rth respondent's ideal point. Rewriting (8), we have for $j = 1, \ldots, p$

$$Y_{jr} = -2\sum_{i=1}^{n}(x_{ri} - \bar{x}_i)\, a_i\theta_{ji} + \text{error}.$$

We then fit the above linear equation using the Y_{jr} as the dependent variable and $\hat{a}_1\hat{\theta}_{j1}, \ldots, \hat{a}_n\hat{\theta}_{jn}$ as the independent variables. As long as $p > n$, a solution can be obtained but the degrees-of-freedom $p-n$ will be small in most applications. Thus, there will be a sizable error in the estimate of a given respondent's ideal point. The estimated ideal points of the respondents should not be trusted, but we can make substantive comments about the average ideal point of a group of respondents if we group the respondents in a meaningful way.

Given the small number of degrees-of-freedom in the ideal point regression, we should reconsider the assumption that all respondents have similar perceptions of all candidates. Although the misperception errors average out in the estimation of the candidate map, there are not enough replications for a given respondent to reduce significantly the error in the ideal point estimate. Since we know that some candidates are better known than others, we use a reduced set of candidates to obtain the ideal points for the 1968 survey -- Nixon, Humphrey, Rockefeller, Johnson, Agnew, and Muskie (Robert Kennedy is eliminated since he had been assassinated several months before the survey).

This completes the formal development of the method. To this point, however, we have not specified how an appropriate value of k is chosen. In applying our method to real data we estimate k simply by trying different values and ascertaining which value yields the closest fit with the internal checks on our assumptions. These checks include: First, all estimated a_{ii}'s should be positive. Second, if $\underset{\sim}{x}_r$ is near $\underset{\sim}{\theta}_j$, the reproduced thermometer score should be near 100. Finally, the distances between $\underset{\sim}{\theta}_j$ and $\bar{\underset{\sim}{x}}$ should correlate significantly with candidate j's average thermometer score. Before we turn to the analysis of artificial and real data, however, in the next section we offer two important caveats to the model.

Two Caveats

Preliminary pretesting of the preceding model reveals at least two contingencies that might necessitate modifying a straightforward application of the procedure (and, in fact, any metric scaling procedure) to real data. Thus, we offer two caveats as warnings about any such application -- warnings that we must consider later when analyzing the election survey data.

The first contingency concerns the possibility that one or more candidates are at a considerable distance from many of the respondents. Suppose, for example, that many citizens regard candidate j as an "extremist" and that relative to other candidates, they prefer to respond with a negative thermometer score for him. Since thermometer scores are necessarily constrained to the bound [0, 100], however, suppose these citizens respond with $T_{jr} = 0$. The problem, then, is that these scores are not represented simply by expression (1), but also by: if $T_{jr} < 0$, set $T_{jr} = 0$. In general, this pulls the candidate closer to the respondents' ideal points than is actually the case, which, in turn, affects the recovery of the positions of other candidates and the estimates of $\underset{\sim}{A}$ in unknown ways. Further, since T_{jr} does not vary in accordance with our assumptions, the covariance between T_{jr} and T_{ir}, $i \neq j$, approaches zero and the method tends to place $\underset{\sim}{\theta}_j$ on a separate dimension.

The simplest means of resolving this problem is to apply the method and check the results for negative a's and dimensions that differentiate only one candidate (or candidate pairs

in the case of running mates). If such a dimension exists, check individual thermometer scores for a disproportionate number of zero scores for the relevant candidate. If there are a great many zeros, rerun the analysis after deleting this candidate from the study.

The second contingency concerns the possibility that the issue space contains one or more of what Stokes terms "valence issues" -- dimensions over which candidate positions but not citizen preferences vary.[16] Suppose, for example, that $\underset{\sim}{\theta}_j = (\theta_{1j}, \theta_{2j}, \theta_{3j})$ but that $\underset{\sim}{x}_r = (x_{r1}, x_{r2}, 0)$ for all respondents r -- i.e., on the third dimension the ideal point of all citizens, x_{r3}, equals zero. Thus, we can no longer assume that there exists a linear transformation of the axes that renders the covaraince matrix $\underset{\sim}{\Sigma}$ equal to the identity matrix I, but rather that

$$\underset{\sim}{\Sigma} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

i.e., the variance of the x's on one dimension is zero. Thus $\underset{\sim}{Y}'\underset{\sim}{Y}$ is of rank 2 rather than of rank 3 so that $\underset{\sim}{Y}'\underset{\sim}{Y}$ fails to reveal the true dimensionality of the space. Moreover our estimates of the mean preference on dimensions 1 and 2, $\bar{x}_1$, and $\bar{x}_2$, contain the effects of the third dimension, and $\theta_{3j}^2 a_3$ cannot be identified (estimated) if θ_{3j} is different for each candidate.

Our application of the method to the 1968 election data suggests that such a dimension exists there. Specifically, while the factorization reveals one or two primary dimensions, several estimates

of $\underset{\sim}{a}$ are negative while the across candidate correlations between the observed and predicted average thermometer scores are less than satisfactory.[17] We resolve this problem thus: we hypothesize the existence of a valence issue that conforms to candidate familiarity and assume that $\theta_{j3} = 1$ for the remaining candidates. This adds an intercept to the linear equation which is fitted by least-squares in order to generate estimates of the a_i's and the rotation. If these assumptions are correct, then the remaining issue weights should be estimated properly. We consider, of course, several alternative partitions of candidates and choose the partition that yields positive estimates of the a_i's and a high correlation between average predicted and observed thermometer scores.

While this procedure is somewhat ad hoc, there are, in fact, sound theoretical reasons for believing that "familiarity" is an issue and that it should receive special treatment in an analysis of election survey data. Our method assumes that citizens possess well-defined perceptions of each candidate's position. In reality, however, citizens are uncertain about positions so that if they are risk averse, respondents should discount thermometer scores by the degree of uncertainty they associate with a candidate.[18] Presumably, however, they perceive some variation among candidates on this criterion. Hence, the perceived riskiness of candidates contaminates the results. But, to the extent that a qualitative assessment of familiarity measures this perception, we can minimize the extent of the contamination.

Artificial Data

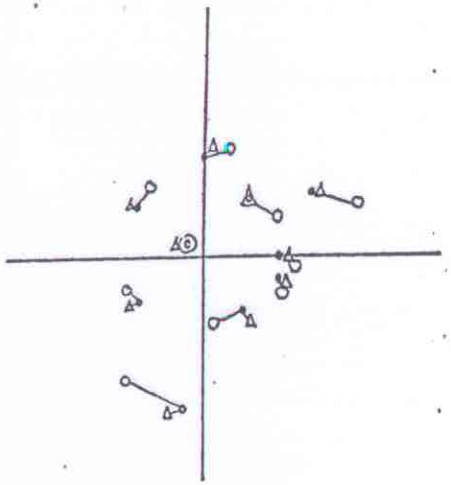
In this section we examine the robustness of the method by applying it to artificial data. In the space allowed, we cannot report all test runs; instead we compare one case in which the data satisfy all assumptions to runs in which alternative key assumptions are violated. We begin by generating a set of two-dimensional preferences using a normal N(0,1) random number generator for a sample size of 1000 respondents. Ten candidates are arbitrarily placed in the space, and, letting $k = 4$ and $a_1/a_2 = .6$ thermometer scores are computed for each respondent. After rescaling these scores so that they fall in the range [0, 100], we then add a normal error term whose standard deviation is 10.95. Each resulting score is then rounded to a member of the set {0, 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 85, 90, 95, 100}. This procedure eliminates scores less than 0 and greater than 100 and produces data that conform to people's apparent tendency to round their responses.

Applying the method to this data -- data that satisfy all assumptions except for rounding -- we assume first that the true value of k (=4) is known. We present the results of that run in Figure 2, where dots denote true candidate positions and circles denote recovered positions. To indicate the effects of rounding we add to Figure 2 (and denote by triangles) the positions we recover when thermometer scores are not rounded. Clearly, rounding distorts somewhat our recovery of spatial locations but, overall, the distortion does not appear significant. One measure of its significance is the ratio $|\hat{\theta}_j| / |\theta_j|$ and, in particular, the maximum and minimum

observed for this ratio. With respect to the preceeding data, all such ratios fall in the range (.88, 1.2) -- i.e., we observe <u>at most</u> a 20 percent error. Our estimate of a_1/a_2, moreover, .632, is surprisingly close to its true value .60, while the correlation, R^2, between the candidate's average thermometer scores and their distances from the mean preference is .9995. Overall, then, the method appears to be quite satisfactory, at least if all assumptions are satisfied and k is known.

Suppose, however, that k is not known. Specifically, in Figures 3, 4, and 5 we contrast recovered and actual candidate positions when $k = 4$ but is mistakenly assumed to equal 1, 2, and 8 respectively. Comparing these figures to Figure 2 we observe some deterioration in our estimation, but this deterioration does not appear to be substantial, i.e., the range of $|\hat{\theta}_j| / |\theta_j|$ becomes (.80, 1.34), or a maximum error of 35 percent. The accuracy of our estimates of $\hat{a}_1/\hat{a}_2$ also decline somewhat (equaling .643, .738, and .570 for $k = 1,2$, and 8 respectively) although again this decline is not disturbing. More important, however, R^2 declines as well (equaling .935, .994, and .996). Thus, if we use this criterion to select an appropriate value for k, we would in fact choose the correct value. Overall, then, the procedure works well if all assumptions are satisfied but when k must be estimated.

Despite these positive results, no statistical methodology can be wholly insensitive to violations of its assumptions. For an example, consider the assumption that $\underset{\sim}{A}$ is diagonal, i.e., that there exists a linear transformation of the axes that renders A diagonal and and $\underset{\sim}{\Sigma} = \underset{\sim}{I}$. Suppose instead that with A diagonal,



igure 2: k=4 and assumed to be known

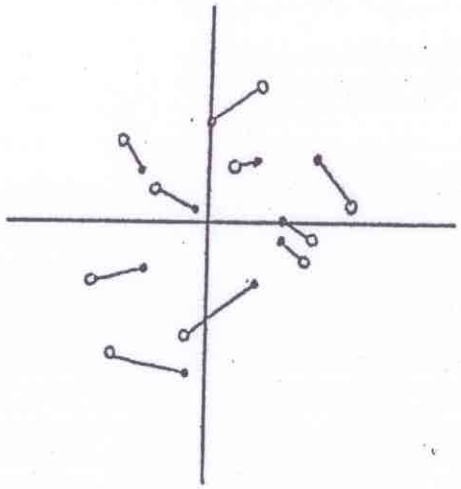
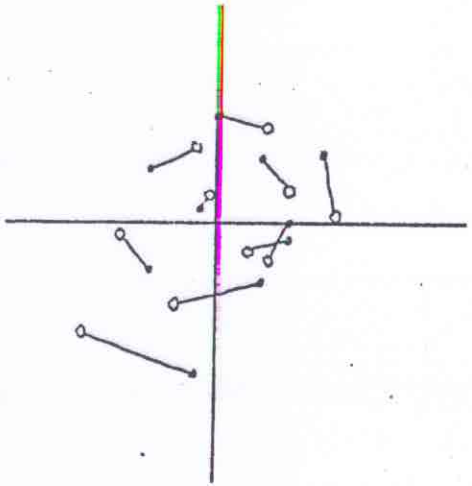


Figure 3: k=4 but assumed to be 1



igure 4: k=4 but assumed to be 2

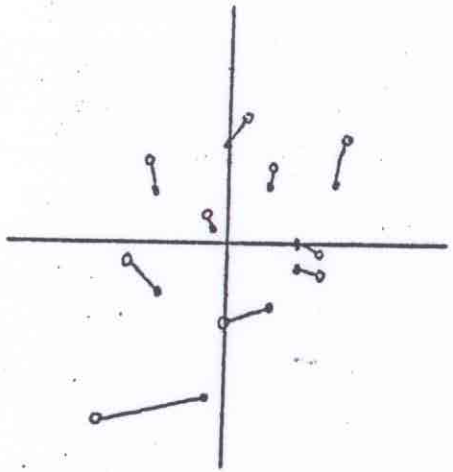


Figure 5: k=4 but assumed to be 8

$$\underset{\sim}{\Sigma} = \begin{bmatrix} 1.0 & \rho \\ \rho & 1.0 \end{bmatrix}$$

where $\rho \neq 0$. Letting $a_2/a_1 = 3.0$, in Figure 6 we contrast the candidate's true and recovered positions for $\rho = .64$. Clearly, correlation in $\underset{\sim}{\Sigma}$ occasions considerable distortion of the recovery (note, though, that a substantial amount of the distortion can be attributed to rotational error). Further our estimate of a_2/a_1 fares badly -- equalling 7.11 as against the true value of 3.0.

The final artifical data run we report here anticipates the problems we encounter later with valence issues. First, we choose a new two-dimensional configuration of candidates that is not dissimilar from the configuration obtained in the next section. Second, we generate a sample of artificial thermometer response scores as before after letting $a_1/a_2 = 2.33$ and after assuming that there exists a valence issue such that $\theta_{j3} = 1$ for some candidates, $\theta_{j3} = 0$ for others, $x_{r3} = 0$, and $a_3/a_2 = 1.00$. The results now of applying the procedure to this data knowing k (=4) but without a correction for the valence issue yields the negative ratio $\hat{a}_1/\hat{a}_2 = -3.0$ and considerable distortion in our estimates of $\underset{\sim}{\theta}_j$ and the mean. While we cannot use the distortion in $\overline{\underset{\sim}{x}}$ and $\underset{\sim}{\theta}_j$ as a test for valence issues (since we cannot observe the true values of $\overline{\underset{\sim}{x}}$ and $\underset{\sim}{\theta}_j$), a negative estimate for an issue weight is a clear warning. Figure 7 shows, in fact, that even if we overlay candidate 1's recovered and true positions, our recovery of the remaining candidate positions (dots denote actual positions and triangles denote recovered positions) is unsatisfactory.

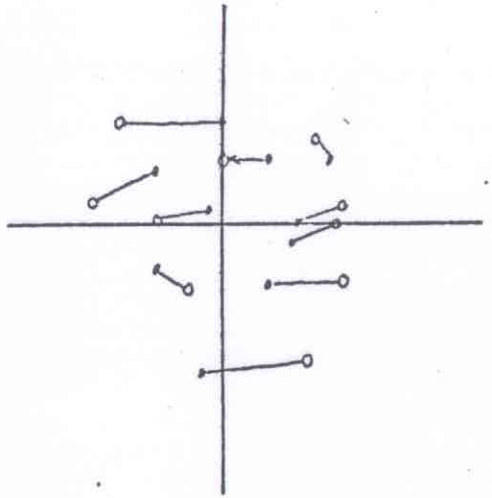


Figure 6: Correlation in Σ

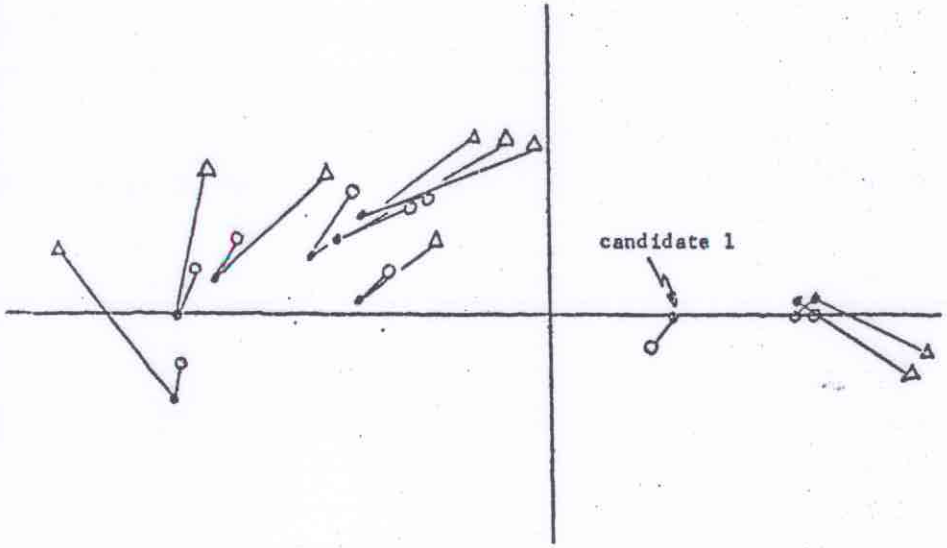


Figure 7: Estimated Candidate Positions with Valence Issue -- Corrected (o), Uncorrected (Δ)

Suppose, however, that we can properly identify the positions of candidates on the valence issue and rerun the procedure accordingly. Now, $\hat{a}_1/\hat{a}_2$ equals 1.79 while $\hat{a}_3/\hat{a}_2$ equals 1.11 (as against the true values of 2.33 and 1.00). Further, as Figure 7 shows, the recovery of candidate positions (denoted by circles) is improved considerably (without any correction for the origin of the space). We conclude, then, that perhaps even more than off diagonal elements in $\underset{\sim}{\Sigma}$ or misidentification of k, valence issues can render the results of our procedure (and presumably any similar procedure that uses thermometer data) distorted or meaningless. But, if such an issue exists and if we can independently estimate or guess the positions of the candidates on it, the procedure can be modified appropriately.

Election Data

We know, of course, that no statistical procedure can be applied indiscriminately to data. The appropriateness of a particular procedure -- such as linear regression analysis -- depends on our research interest and on our willingness to assume that the data conform to the statistical and structural assumptions of the procedure. Thus, beyond examining artificial data, we cannot "test" the methodology developed in the preceeding sections in the sense that we test a theory. Instead, we must apply it to data that we assume are appropriate and assess the method's adequacy in more intuitive ways. First, we require that the resultant spatial maps make some sense. That is, if our estimates of the candidates' positions depart significantly from intuitive presuppositions or from the estimates

others obtain using different techniques, then we must either formulate an explicit hypothesis about why those techniques are inappropriate or consider the hypothesis that our assumptions do not adequately approximate the data's properties. Second, at least two internal checks on our assumptions must be satisfied: (1) all estimated a's must be positive, and; (2) the distances between $\bar{x}$ and the $\hat{\theta}_j$'s should correlate across candidates with the respective candidate's average thermometer score across respondents. Finally, we examine the distribution of nonvoters in the estimated spaces and examine the extent to which it is consistent with one of two hypothesis about the causes of abstention -- alienation or indifference.

We begin by noting that the results of a straightforward application of our procedure to the 1968 SRC election survey data is unsatisfactory in several respects. First, while two dimensions are estimated, the a_i's for one or more of these dimensions are negative. Second, much like the results reported elsewhere, George Wallace and Curtis LeMay are located on a separate dimension, far from other candidates. Third, the distances between the $\hat{\theta}_j$'s and $\bar{x}$ do not correlate highly with the candidates' average thermometer scores across respondents.

A simple examination of the data, however, suggests immediately why a naive application of the procedure is inappropriate. First, the data seem overly noisy despite our allowances for noise in the model. In particular, far too many respondents assign scores of fifty to all but a few candidates -- suggesting that they either possess no well-defined attitudes or that they fail to respond and

discriminate carefully. Our resolution of this problem is to drop from the sample all respondents who report a score of fifty for four or more candidates (out of 12). In addition, respondents who did not score all the candidates named were removed since the model assumes that the respondent be politically aware. The table in Footnote 19 shows the difference between the total and subsample income and education distributions, and we also give a breakdown of reported voting statistics for the two groups. The fact that we analyze a subsample who have higher income and education than the total does not invalidate the results. We are concerned with testing the implications of spatial theory.

The second problem with the data concerns the Wallace and LeMay responses. Consider Table 1, where we report the number of zero scores given each candidate by the filtered sample's respondents.

Clearly, Wallace and LeMay are distinct from all other candidates in this respect. While the respondents giving zero scores to these two running mates number in the hundreds, their closest "competitor" is Reagan with a total of fifty-six zero scores. This suggests that a significant percentage of the response scores for these two candidates do not conform to our model's assumptions. In accordance with our first caveat, then, we do not include Wallace or LeMay in our analysis of the data.

A third potential problem concerns the assumption that all respondents share a common $\underset{\sim}{A}$-matrix. We know, of course, that such an assumption is probably false with respect to "actionable issues" such as Vietnam, bussing, or inflation. Our procedure, however, seeks to recover underlying dimensions of taste that, in

Table 1: Number of Respondents Reporting Zero Thermometer Scores by Candidate, 1968

	Full Sample	Democrats	Independents	Republicans
Wallace	265 (43%)	142	73	97
LeMay	181	107	46	59
Reagan	56	44	18	7
McCarthy	44	19	14	21
Johnson	43	11	16	21
Romney	42	22	16	13
Agnew	41	35	12	3
Humphrey	37	7	17	24
Rockefeller	34	16	11	14
Kennedy	32	12	15	16
Muskie	23	9	8	12
Nixon	16	15	3	0
N	620	317	206	244

Note: Due to the overlap in these groups the number of Democrats, Independents, and Republicans exceed 620.

conjunction with socio-economic circumstances among other things, presumably determine a person's preferences on actionable issues. There is little if any evidence to suggest that the relative importance of these dimensions varies systematically across the electorate (if only because methods for discussing such variations do not exist). Nevertheless, as a partial control for variations in $\underset{\sim}{A}$, we divide the sample into three subsamples -- Democratic and Republican party identifiers and Independents.[20] Clearly, other partitions should be tried as well -- perhaps along socioeconomic characteristics. The party identification partition, nevertheless, should provide some minimal control over possible variations in $\underset{\sim}{A}$. Furthermore, since, across subsamples, we do not anticipate significant variations in k, in the number and relative ranking of the issues, and in relative candidate position, this model also provides a check on the model's sensitivity to purely stochastic differences.

Applying the model to this data, our first task is to estimate the dimensionality of the space. Table 2 reports the percent variance explained by the first five eigenvalues of the matrices $\underset{\sim}{Y}'\underset{\sim}{Y}$ and $\underset{\sim}{W}$ for each of the three groups for the two-dimensional solution. The same statistics are shown for $\underset{\sim}{W}$ in the one-dimensional solution.[21] A consideration of these eigenvalues indicates a choice of one or two dimensions for the Democrats and Republicans, and perhaps as many as three for the Independents. From a detailed analysis of the data it is clear that the Independents are a heterogenous group and that they exhibit the most perceptual variation. Therefore, we base the determination of the dimensionality

of the candidate space for the Independents upon the results for the other two groups.

Table 3, however, shows the ordering of the candidates if we extract only one dimension.[22] Although this ordering conforms in an intuitively satisfying way to a party identification or left-right breakdown, the results are less than satisfactory.[23] Specifically, we estimate the mean citizen preference to be far from all candidates, which is unreasonable. If we proceeded directly to a two-dimensional recovery, however, the results remain unsatisfactory. While the relative positions of candidates and the order of differences between pairs are nearly identical to those Rabinowitz obtains from his modified algorithn, a negative axis weight is estimated for Independents, and among Democrats and Independents, virtually all variation in the candidate's positions occurs along a single dimension.[24] Our hypothesis is that at least one valence issue is distorting our results.

TABLE 2

EIGENVALUES OF $\underset{\sim}{X}'\underset{\sim}{X}$ AND $\hat{\underset{\sim}{W}}$ USED TO DETERMINE DIMENSIONALITY OF THE SPACE

	Democrats	Independents	Republicans
$\underset{\sim}{X}'\underset{\sim}{X}$	53.6%	53.5%	45.0%
	12.9	11.3	14.4
	8.2	9.8	8.9
	6.5	6.4	6.9
	4.9	4.7	5.9
$\hat{\underset{\sim}{W}}$ Two-dimensional Solution	84.9%	89.7%	83.2%
	15.1	10.3	16.8
	5.7	10.1	7.7
	3.0	2.7	4.3
	2.7	0.5	2.7
$\hat{\underset{\sim}{W}}$ One-dimensional Solution	100.0%	100.0%	100.0%
	16.6	10.4	17.6
	5.3	10.0	7.4
	2.1	2.2	3.4
	1.2	-1.0	1.0

Table 3: Ordering of Candidates on Principal Dimension with No Valence Issue

Republicans	Democrats	Independents
Humphrey	Humphrey	Johnson
Kennedy	Johnson	Kennedy
McCarthy	Kennedy	Humphrey
Romney	Muskie	Romney
Johnson	Rockefeller	Rockefeller
Rockefeller	McCarthy	McCarthy
Muskie	Romney	Muskie
Reagan	Nixon	Nixon
Nixon	Reagan	Agnew
Agnew	Agnew	Reagan

Recall that there are sound theoretical reasons for supposing that if variations in the riskiness or familiarity of candidates is a salient concern, this concern should be given special treatment. Consequently, we hypothesize that respondents discount the thermometer scores of unfamiliar candidates. One problem with integrating this hypothesis into our analysis is the selection of an appropriate partition of the candidates. Here, we use some intuition and require, additionally, that all a's be positive and that the partition increase the across candidate correlation between the observed and predicted average thermometer score. While we cannot consider all possibilities, the partition that works best among those examined is:

$\theta_{j3}=0$	$\theta_{j3}=1$
Johnson	Muskie
Nixon	Agnew
Kennedy	Reagan
Humphrey	Rockefeller
	McCarthy
	Romney

The value of k which works best is 4, which is to say that respondents fail to discriminate sharply between candidates who are far from their ideal point. A review of the data reveals that the Democrats do not distinguish as well among the Republicans as they do among the candidates of their own party, and similarly for the Republicans. It was this observation which led us to the development of the method for $k=4$. The analysis using $k=2$ (straight distance) is almost as good, but squared distance is significantly poorer.

Using this assumption in our regressions to estimate $\underset{\sim}{\theta}$, $\underset{\sim}{A}$, $\underset{\sim}{\Gamma}$, and $\bar{\underset{\sim}{x}}$, Figures 8, 9, and 10 show the resultant spatial maps (the symbol "0" denotes the estimated mean population preference, $\bar{\underset{\sim}{x}}$), while Table 4 provides several related statistics.[25]

Table 4: Best Fit Predictions of the 1968 Vote and Related Statistics

	Republicans	Democrats	Independents
Predicted Vote			
Nixon	96.8	17.9	62.7
Humphrey	3.2	82.1	37.3
Actual Vote			
Nixon	92.6	21.6	62.7
Humphrey	7.4	78.4	37.3
% Vote Correctly Predicted	95.9	88.3	84.9
$\hat{a}_2/\hat{a}_1$	1.74	5.43	2.44
$\hat{a}_3/\hat{a}_1$	.79	.39	1.12
R^2	.995	.972	.983
k	4	4	4

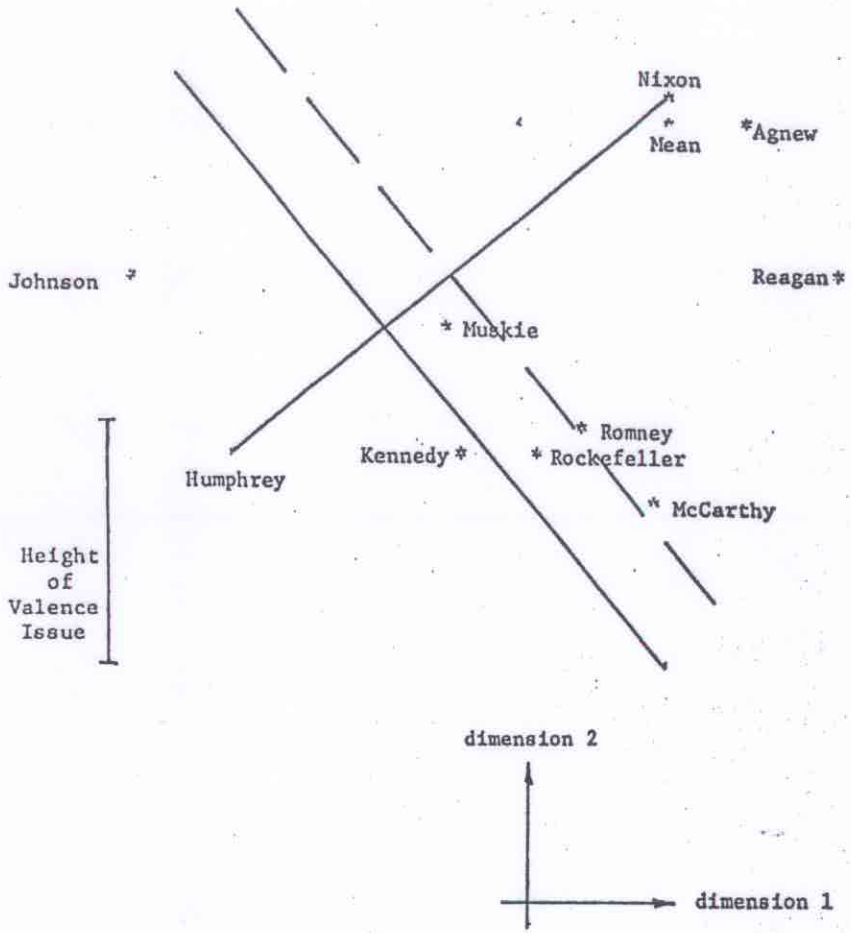


FIGURE 8: Republicans

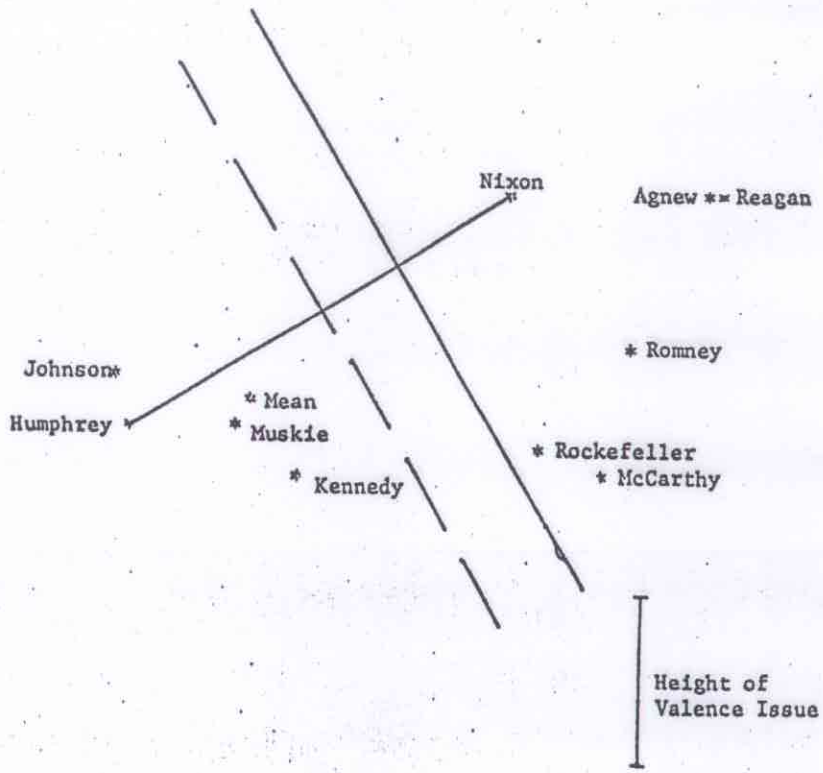


FIGURE 9: Democrats

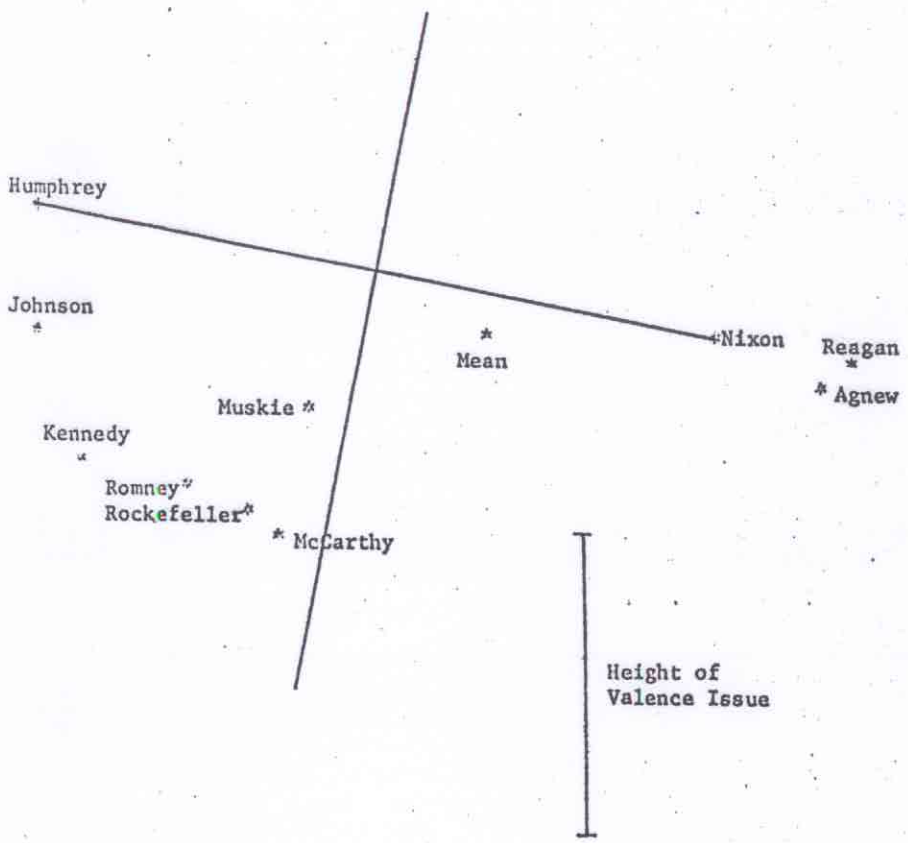


FIGURE 10: Independents

With respect to Table 4, the most important effect of including the valence issue assumption in our regressions is the increase in the relative salience of, and variability of the candidates over the vertical dimension. Aside from differences that we can attribute to rotation, the ordering of the candidates along the horizontal dimension conforms to Table 3 (or, more precisely, to the orderings in note 20). The vertical dimension, then, is the dimension that previously failed to discriminate candidates. Aside from this change, several additional aspects of these results warrant emphasis. First, note that aside from variations in distance, the candidates relative positions are quite similar in all three figures. Second, the estimated mean ideal points conform to our expectations, i.e., the Republican mean preference is near Nixon, the Democratic mean is near Muskie and Humphrey, and the Independent mean lies midway between Muskie and Nixon. We note also that the inclusion of the valence dimension decreases the sum of squared errors in our regressions by a factor of two.

The one deviation from a consistent pattern across subpopulations concerns Independents. Specifically, while the order of the candidates along the vertical dimension for Democrats and Republicans places Nixon, Agnew and Reagan at one extreme, McCarthy and Kennedy at the other and Jonson, Humphrey and Muskie near the center, for Independents Humphrey is at one extreme, followed by Johnson and Nixon. Observe, however, that a 30° counter-clockwise rotation of the Independents' space brings that space in line with Figures 8 and 9. Given the quality of the data, a 30° error in $\hat{\Gamma}$

seems well within tolerances -- and is, in fact, consistent with the rotation we observed in Figure 7 using artificial data.

It is interesting at this point to compare the variance of candidate positions to the variance of citizen ideal points. The pattern is the same for all three groups and so, in Figure 11, we portray the distribution of Republican ideal points, letting "*'s" denote candidate position (numbers and letters are used to denote the number of ideal points at a particular point on the grid such that "1" = 1,..., "A" = 10, "B" = 11 and so on). Clearly, the dispersion of the x_r's is far greater than of the θ_j's.

With respect to predicting votes, the dotted lines in Figures 8, 9, and 10 represent our theoretical predictor as to how citizens divide between Nixon and Humphrey.[26] That is, all voters with ideal points to the left of this line are closer to Humphrey than to Nixon, and should, therefore, vote for Humphrey. All voters to the right of the line should vote for Nixon. The dark solid line corresponds to the line parallel to the bisector that maximizes the percent of the votes correctly predicted We use this line in reporting the percentages of the votes for each candidate in Table 4. For Independents, the two lines are identical. For Republicans, the line is biased toward Nixon; and for Democrats, it is biased toward Humphrey. There are at least two explanations for this "bias."

The first explanation is statistical and does not relate to the model directly. However, it is seen most easily by utilizing the model. Consider, again, the predicted votes for the Democrats. If we predict votes based solely according to which candidate the citizen is closer, Nixon or Humphrey, we underestimate Humphrey's votes and

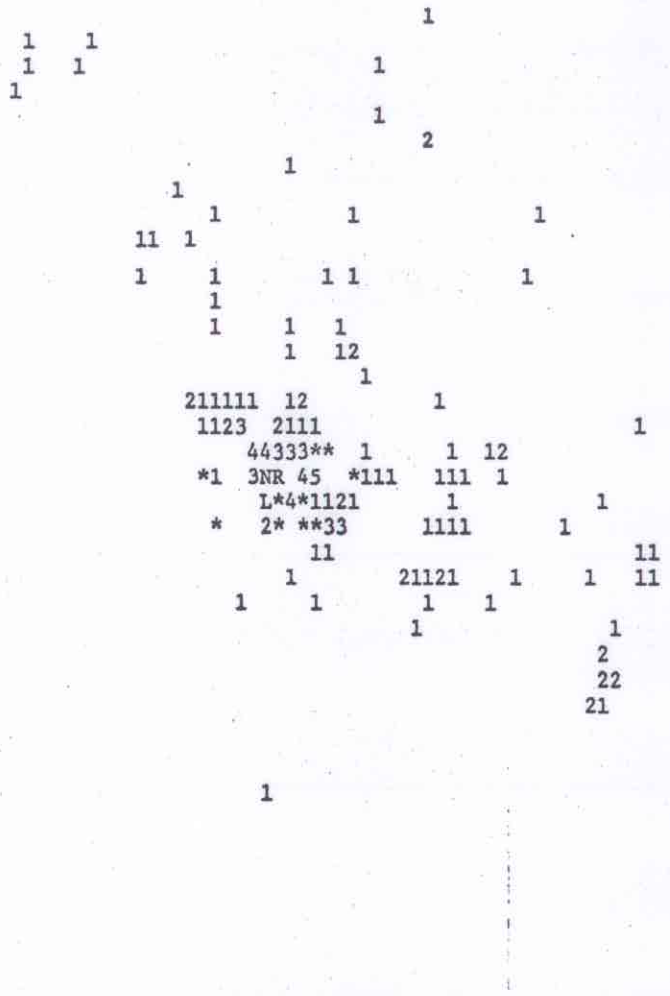


Figure 11: Distribution of Republican Identifiers, with * Denoting Candidate Position

overestimate Nixon's votes. This occurs because we utilize the estimated citizen positions to predict votes. Let us illustrate this effect by the following one-dimensional situation. Nixon and Humphrey are the only two candidates. We are considering the behavior of the Democrats, so we assume that the majority of the citizens are located nearer Humphrey. This situation is illustrated by Figure 12, where $f(x)$ represents the true distribution of the ideal points in the population. We utilize $\hat{x}_i = x_i + \varepsilon_i$ to predict voting behavior, and the distribution of $\hat{x}_i$ is represented in the figure by $f(\hat{x})$. The probability of voting for Nixon, then, is the area under $f(\hat{x})$ to the right of the vertical line. The probability of predicting a vote for Nixon based upon the estimated citizen position is the area under $f(x)$ to the right of the vertical line. Clearly, the probability of predicting a vote for Nixon is greater than the probability of a vote for Nixon. Thus, predicting votes based on the distance from the candidate, utilizing the estimated citizen positions, results in overestimating Nixon's vote and consequently underestimating Humphrey's vote. The effect is reversed if we predict votes for Republican respondents. Here we overestimate Humphrey's vote and underestimate Nixon's vote. There is no equivalent effect for the Independents, since the votes are more evenly distributed between Nixon and Humphrey.

This same reasoning shows that we increase the probability of correctly predicting votes if we move the line dividing the voters toward Nixon for the Democrats and away from Nixon for the Republicans. From Figures 8 and 9, it can be seen that this is exactly what has been done in order to correctly predict the maximum number of votes.

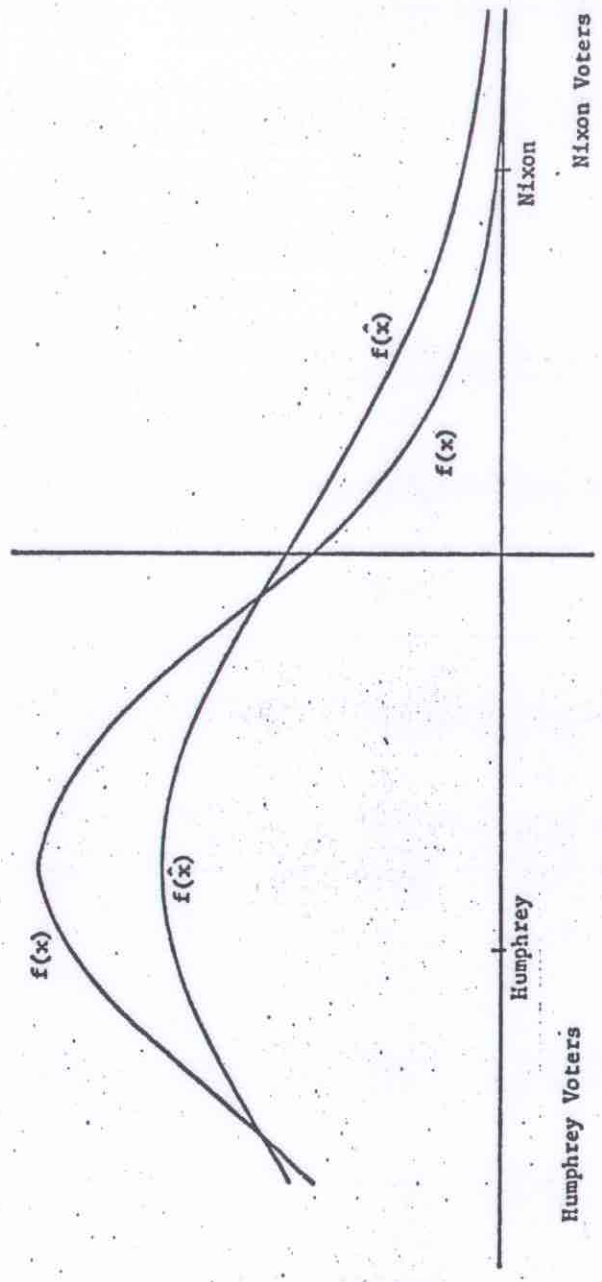


FIGURE 12
Probability Density Functions of the Ideal Points in the Population and the Estimated Ideal Points

In addition to this reasoning, there is a second cause for bias. The thermometer responses were secured after the election, and we suspect that this affects our predictions. Note that for the Democrats, if we move Nixon closer to Agnew and Reagan, the disparity between the two lines diminishes. It is reasonable to suppose that Nixon's scores are enhanced by his new status as President-elect -- he is moved closer to the center of the space than is representative of his correct position.

An important question, now, is whether these spatial maps are meaningful and useful. Aside from commenting that dimension 1 resembles a "left-right" continuum and that 2 principally differentiates by identification, our imaginations are not so limited doubtlessly as to preclude the construction of a good story about any observed dimension. There are other means, though, for assessing the substantive content of dimensions.

Included in the 1968 Election Survey are several questions that elicit from the respondent a thermometer score for several political categories, including "Viet-Nam War Protestors," "Liberals," "Republicans," "Democrats," and so on. Rusk and Weisberg use these scores to estimate (independent from candidates) the spatial configuration of these groups. They then overlay this space over the candidate space, to infer the substantive content of the original dimension. We illustrate here an alternative procedure: the scores for each group are used to divide respondents into five groups. Group 1 consists of all respondents who give a score from 0 to 19; Group 2 from 20 to 39; Group 3 from 40 to 59; Group 4 from 60 to 79; and Group 5 consists of respondents who gave scores from 80 to 100. Within each group we then calculate

the mean of the estimated respondents' positions. This, then, yields five average positions that range from greatest dislike towards a category to greatest liking for that category. By plotting the positions of these five groups for each category we can then ascertain the relation between citizen preferences for candidates and issue relevant categories of political actors.

The thermometer issue question is applied to a total of nineteen issue-categories. Not all of these are useful, nor do all of them reveal any relationship to the respondents' positions. In Figure 13 through 17 we show some of the relationships between citizen feelings on the "issues" and their positions relative to the candidate. Table 5 presents the number of respondents in each of the five groups for each of these figures.

The relationship between Republicans and their feelings toward "Viet-Nam War Protestors" is shown in Figure 13. There is no thermometer question by which citizens could express their feelings on Viet-Nam directly, but the striking alignment of the groups on the vertical dimension indicates that the issue of Viet-Nam is related to that dimension. The response of the Democrats to the same question is similar, but not as well defined. The Democrats express their attitude on Viet-Nam more clearly when asked to indicate how they feel toward "The Military." The plot of the respondents' positions on this issue is shown in Figure 14 and again we see a strong vertical alignment of the five groups.

The true nature of the vertical dimension is perhaps best indicated, though, by the respondents' feelings toward "Liberals." The

TABLE 5

NUMBER OF CITIZENS IN EACH GROUP IN FOLLOWING FIGURES

Figure	13	14	15	16	17	18	19
Party Group	Rep	Dem	Dem	Rep	Ind	Dem	Rep
Issue	Viet-Nam War Protestors	Military	Liberals	Liberals	Liberals	Republicans	Democrats
Group 1: scores 0-19	117	11	17	34	23	21	25
Group 2: scores 20-39	36	11	18	26	14	14	15
Group 3: scores 40-60	74	74	79	137	112	185	135
Group 4: scores 61-80	4	58	49	30	27	50	43
Group 5: scores 81-100	9	161	51	14		45	23

Note: A score of 100 indicates a great liking for an "issue," while a score of 0 indicates a great dislike for the "issue."

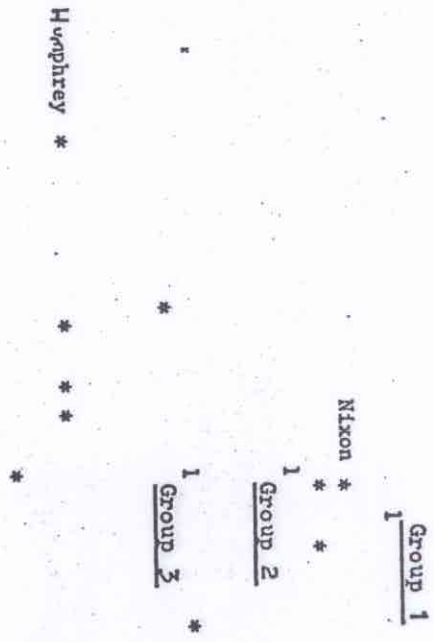


FIGURE 13

Republicans: "How Do You Feel About Viet-Nam War Protestors?"*

*"Groups 4 and 5" are deleted since too few respondents are of this type.

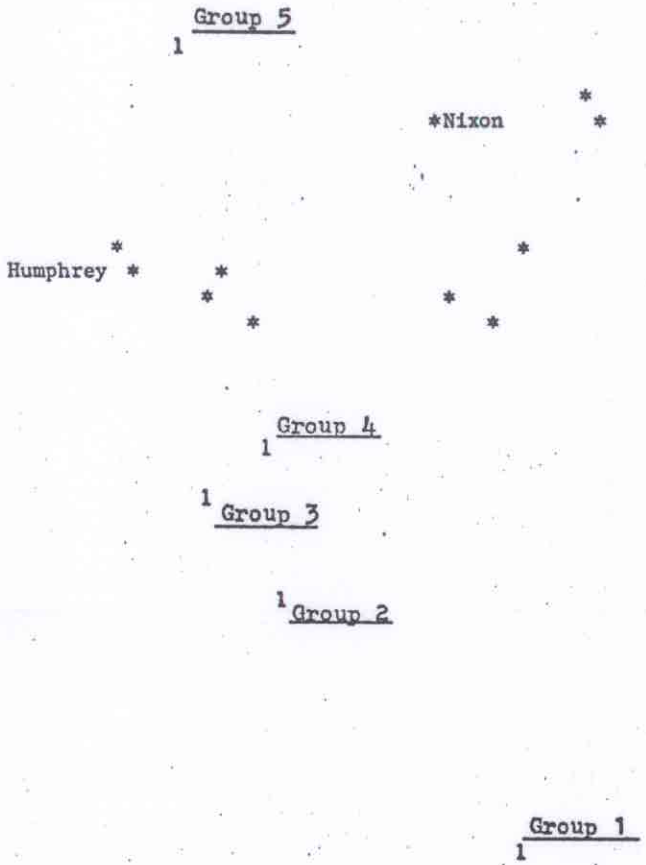


FIGURE 14

Democrats: "How Do You Feel About the Military?"

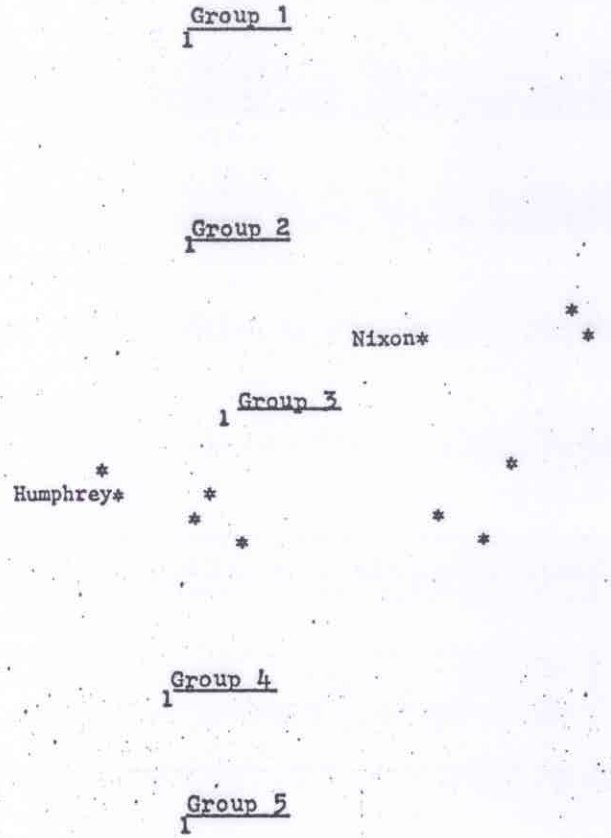


FIGURE 15

Democrats: "How Do You Feel About Liberals?"

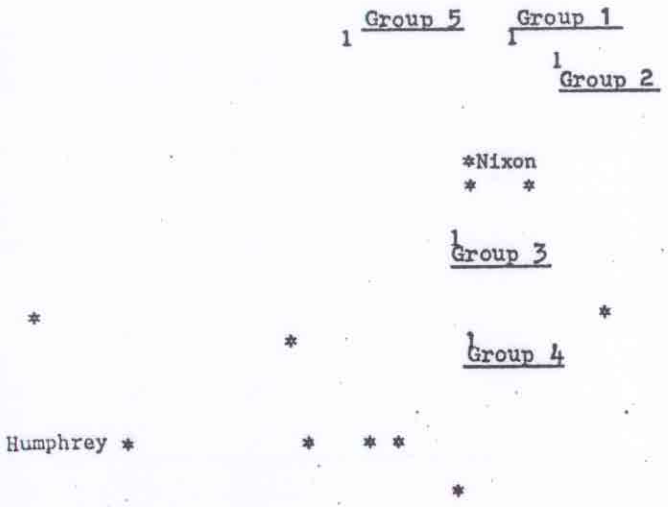


FIGURE 16

Republicans: "How Do You Feel About Liberals?"

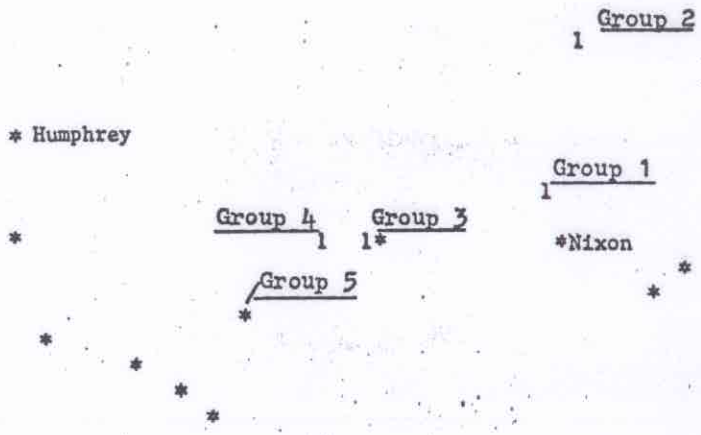


FIGURE 17

Independents: "How Do You Feel About Liberals?"

plot of citizen preference groupings is shown in Figures 15, 16, and 17. Within the Democratic group, the vertical dimension is clearly related to this dimension. Thus we see that the vertical dimensions in fact represent a broad range of social issues that can be interpreted in "liberal-conservative" terms. The Republican response to this question (see Figure 16) is not as well defined. If we combine Groups 1 and 2, and combine Groups 3, 4, and 5, however, we see that Republicans as well view the vertical dimension in the same way as Democrats. The response of Independents (see Figure 17) reveals the effects of the rotation of their candidate space relative to the space we estimate for Democrats and Republicans. If we combine Groups 1 and 2, and combine Groups 3, 4, and 5, we see that the "liberal" alignment is rotated from the direction we observe for the Democrats and Republicans. This agrees with our earlier observation that the estimated candidate configuration for Independents disagrees with that of the Democrats and of the Republicans only up to a rotational error.

Other similar vertical alignments of groups occurs with respect to nearly all remaining thermometer questions. Of all the questions asked, only one differentiates citizens on the second dimension -- party identification. Yet even here there is some correlation with citizen positions on the liberal-conservative dimension. Figure 18 indicates the feeling of Democrats towards "Republicans." Combining Groups 1, 2, and 3, and combining Groups 4 and 5, we see an alignment with the line connecting Nixon and Humphrey that intersects the vertical dimension at an approximately 45° angle. A similar pattern is observed in Figure 19 where we represent the feelings of Republicans

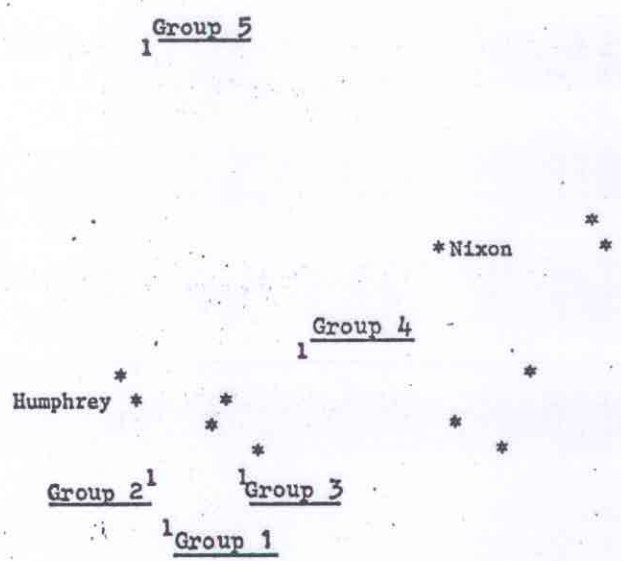


FIGURE 18

Democrats: "How Do You Feel About Republicans?"

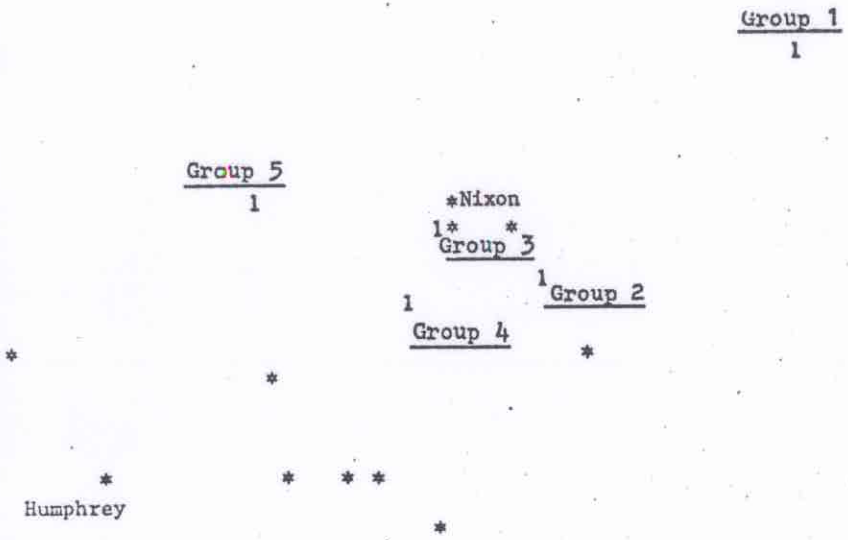


FIGURE 19

Republicans: "How Do You Feel About Democrats?"

for "Democrats." Here we can combine Groups 4 and 5 and combine Groups 2 and 3 to clarify the effect.

This, of course, does not exhaust the possibilities for "issue analysis," i.e., the mean by which we can attempt to secure a substantive interpretation of the recovered dimensions. Attitudes elicited on other questions in the survey can be used as well. For example, in a preliminary analysis of the 1972 Election Survey, we find that the thermometer scores on the inflation issue line up on the horizontal axis, which supports the conjecture that the party axis involves economic issues. We conclude that the dimensions in Figures 8, 9, and 10 "make sense" and reveal that citizens do not evaluate candidates simply in terms of party identification or in terms of ideology; instead, these two notions both operate (albeit in a correlated fashion with respect to the distribution of preferences) to form a citizen's evaluation of a candidate.

Conclusions

No statistical methodology can yield meaningful results if the data do not satisfy its assumptions. Consequently, much additional empirical research remains -- if only to ascertain whether our method is appropriate for available data and whether alternative treatments of the data can resolve apparent or potential problems. Alternative filters and partitions of sample populations might be considered, for example, as well as "valence issues" in which more than two candidate positions are admitted. Additionally, following Rusk and Weisberg, we might use thermometer responses for groups such as the "military," "police," "liberals," etc. in conjunction with those for candidates to identify better the substantive content of the recovered dimensions and to ascertain whether dimensions exist that we cannot recover using candidate stimuli alone (because the candidates' positions do not vary on them). The analysis in the preceeding section, however, is intended simply to illustrate an application of the method, to provide an intuitive and qualitative test of its adequacy, and to highlight tentative resolutions to several problems.

Theoretically, the method should be refined in several respects -- including the development of statistical tests to check which of its internal assumptions the data violate, such as non-diagonal $\underset{\sim}{A}$ and correlated errors. At this point, we can only guess at the cause of such problems as negative a's, low correlations, and rotational ambiguities across subpopulations and, in fact, we can detect these problems only after considerable "playing" with the data.

Aside from the usual "the past is prologue . . . " routine, let us conclude by emphasizing an important problem that we otherwise bypass in this essay. That problem concerns basic vs. actionable issues. Doubtlessly, if we were to sit down and list all contemporary substantive issues or criteria citizens appear to use in evaluating candidates -- such as, the candidates' positions on bussing, civil rights, minimum wage scales, ecology, public transportation, personality, age, party identification, tariffs, farm price supports, tax policy, and education policy -- we would conclude that the election strategy space contains many more issues than the two or three dimensions we recover here. Our resolution of this apparent disparity is the argument that there exist two spaces: The first contains all actionable issues, i.e., all substantive criteria one or more citizens use for evaluating candidates. In this space, of course, the assumption that all citizens share a common salency matrix is untenable since farmers are more likely to be concerned with farm price supports than with bussing while the converse is true for blacks. The second space, however, is assumed to consist of the more general and common evaluative criteria that determine people's preferences over substantive issues. It is this space, presumably, that our method recovers. If the existence of two such spaces constitutes a reasonable conceptualization of preferences, however, one problem is whether our method in fact recovers the basic space or some combination of the two (i.e., are two or more substantive issues collapsing imperfectly into one basic dimension). A second, and more interesting problem is, even if the conceptualization is useful, how do we infer substantive issues from basic dimensions.

This latter problem is a practical concern of considerable importance. Specifically, if we wish to evaluate a candidate's campaign -- a campaign that consists of adopting alternative positions on substantive issues -- then our method is relevant only if we can ascertain the relationships between positions on actionable and basic issue dimensions. After looking at the maps, for example, we might advise Humphrey to move to the right on the horizontal dimension, but how do we translate this advice into a prescription of the positions he should adopt on, e.g., bussing and equal opportunity legislation.

Clearly, then, our method at best provides only a partial resolution of the problems one encounters when attempting to render spatial models of party competition more empirically relevant. Our hope, however, is that by developing such methods -- methods grounded in the basic assumptions of spatial theory -- and by applying them to data, we can more rigorously ascertain the inadequacies of those models and modify them accordingly.

*This research was supported in part by the National Science Foundation.

1. Many spatial models, of course, do not use the weighted Euclidean metric assumption, but models that employ this assumption are the most well developed. For the use of other metrics see R. E. Wendell and S. J. Thorsen, "Some Generalizations of Social Decisions Under Majority Rule," Econometrica, forthcoming, September, 1974, and Douglas Rae and Michael Taylor, "Decision Rules and Policy Outcomes," British Journal of Political Science, 1 (January, 1971), pp. 71-90.

2. See William H. Riker and Peter C. Ordeshook, An Introduction to Positive Political Theory (Englewood Cliffs: Prentice Hall, (1973), cp. 11, 13, and; Richard McKelvey and Peter Ordeshook, "Symmetric Spatial Games Without Majority Rule Equilibria," American Political Science Review, forthcoming.

3. Examples of this research include Howard Rosenthal and Subrata Sen, "Electoral Participation in the French Fifth Republic," American Political Science Review, 67 (March, 1973), 29-54, and; Benjamin I. Page and Richard A. Brody, Indifference Alienation and Rational Decisions: The Effects of Candidate Evaluations on Turnout and the Vote, Public Choice (Summer, 1973).

4. Jerrold G. Rusk and Herbert Weisberg, "Perceptions of Presidential Candidates," Midwest Journal of Political Science, 16 (August, 1972), 388-410; Weisberg and Rusk, "Dimensions of Candidate

Evaluation," American Political Science Review, 64 (December, 1970), 1167-85; Hans Daalder and Jerrold G. Rusk, "Perceptions of Party in the Dutch Parliament," in S. C. Patterson and J. C. Wahlke, eds., Comparative Legislative Behavior (N.Y.: Wiley, 1972), 143-98, and; George B. Rabinowitz, Spatial Models of Electoral Choice (Ph.D. dissertation, University of Michigan). See also Gary A. Mauser, "A Structural Approach to Predicting Patterns of Electoral Substitution" in Romney, Shepard, and Nerlove, eds. Multidimensional Scaling, Vol. 1 (N.Y.: Seminar Press, 1972).

5. Briefly, Rusk and Weisberg use the correlations between scores to rank pairs of candidates by "closeness." This ranking of pairs is then used with the nonmetric Shepard-Kruskal scaling routine to obtain a spatial configuration. Rabinowitz observes, though, that these correlations are sensitive to the distribution of citizens in the space. Rabinowitz's algorithm calculates for each citizen the sum and differences in the thermometer scores for each pair. The average of the maximum difference and minimum sum is used to obtain a measure of the "distance" between pairs. These distances are then used to rank pairs for application of the Shepard-Kruskal routine. Rabinowitz proves, moreover, that if citizens thermometer scores are a monotonic function of any Minkowski metric, then, with error free data, his algorithm is accurate. With noisy data the algorithm is more complicated, but it tends to push the candidates farther

apart than they actually are. The statistical effects of noise on the final estimate of intercandidate distances have not been traced, but it is evident that the error in estimation does not decrease necessarily with increasing sample size.

6. Warren S. Torgerson, Theory and Methods of Scaling (N.Y.: Wiley, 1958), c.p. 11; and; Peter H. Schoneman, "On Metric Multidimensional Unfolding," Psychometrica, 35 (1970), 349-366. On the problems with directly factor-analyzing the thermometer data see C. H. Coombs and R. C. Kao, "On a Connection Between Factor Analysis and Multidimensional Unfolding," Psychometrica, 25 (1960), 219-231, and; N. Cliff and J. Ross, "A Generalization of the Interpoint Distance Model," Psychometrica, 29 (1964), 167-176. Briefly, Coombs and Kao observe that with p candidates and r dimensions, $p > r$, $r + 1$ dimensions are recovered, but they suggest that the extra dimension can be attributed simply to the $\underset{\sim}{x}'\underset{\sim}{x}$ term in expression (1). The problem, however, aside from developing an adequate treatment of error, is that the rotational ambiguity of factor analysis renders it difficult if not impossible to sort this dimension from the others.

7. For a more extensive discussion of these problems see Kenneth Shepsle "Uncertainty and Electoral Competition," paper presented at Mathematical Collective Decisions Conference, Hilton Head, S. C., 1971. In the context of our analysis there is then an ambiguity between utility scales and the cardinality of dimensions.

Specifically, we use utility (or equivalently, an appropriate distance metric) to represent individual sensitivities to distance measured in a common metric space.

8. Op. cit. Aside from the obvious sources of error, note that we assume that all respondents abide by the identical 0-100 scale whereas, for example, three citizens with identical preferences and perceptions might in fact score the same candidate 10, 20, and 30 (i.e., different voters weight distance differently). Looking at the raw data it does in fact appear that some citizens respond in the 30-70 range only while others respond in, say, the 10-90 range. More formally, citizen r's thermometer responses may be related thus to distance

$$(50-D_{jr})b_j + 50$$

where b_j represents the idiosyncratic features of the citizen's scale. Suppose, however, that these idiosyncracies are not related to ideal point positions. Then, we can model b_j as an error term with mean 1 that is uncorrelated with D_{jr}. (Note that setting $b_j=1$ yields the original model). Letting $c_j = b_j-1$ gives

$$(100-D_{jr}) + c_j(50-D_{jr})$$

where $c_j(50-D_{jr})$ is simply an additive error term with mean 0 and uncorrelated with D_{jr}, in conformity with our assumptions about ε_{jr}.

9. At this point the similarities between this analysis and those of Torgerson, op. cit., and of Schoneman, op. cit., should be noted. Briefly, Torgerson assumes data in the form of interpoint distances between the stimuli; thus, his observations are of the form $D_{ij} = (\underset{\sim}{\theta}_i-\underset{\sim}{\theta}_j)'(\underset{\sim}{\theta}_i-\underset{\sim}{\theta}_j)$. By subtracting the means over i and j and by factoring the resulting adjusted distance matrix, he is able to estimate $\underset{\sim}{\theta}$ up to an unknown rotation. Schoneman applies Torgerson's idea to the present model. By subtracting the means over j and over r he obtains $\underset{\sim}{\theta}'\underset{\sim}{x}_j$ so that by factoring the resultant $n \times p$ matrix and solving a system of simultaneous equations, he estimates $\underset{\sim}{\theta}_r$ and $\underset{\sim}{x}_j$ up to some unknown rotation. Schoneman applies the technique to artifical data generated with error and he realizes that, while his model does not formally treat error, the resultant simultaneous equations are a regression model. He applies OLS but an analysis of error reveals that the equations constitute an error in variables regression model. Thus, OLS yields biased estimates of the parameters and the magnitude of this bias is invariant with sample size. Further, even with prior knowledge of the exact error distribution, it is difficult if not impossible to correct the OLS regression estimates.

10. For a general statement and proof of Result 1 see Lawrence Cahoon "Locating a Set of Points using Range Information Only," PhD Thesis, Statistics, Carnegie-Mellon University, 1975.

1. Experience with artificial data for $n=2$ has shown that better estimates of the parameters are obtained if we search for $\hat{\sigma}^2$ which makes the average of the last $p-2$ eigenvalues be equal to zero.

2. A brief note is in order at this point regarding the identifiability of $\underset{\sim}{\Gamma}$. Multiplying any row or column of $\underset{\sim}{\Gamma}$ by -1 results in an orthogonal matrix. Also, interchanging columsn of $\underset{\sim}{\Gamma}$ also results in an orthogonal matrix. Because of this, the points $\underset{\sim}{\theta}_j$ are not identified exactly. The form of the distance observations results in this identification problem. A reflection of the candidate and citizen position across either axis does not change the set of distance observations. A rotation of the points by 90°, together with interchanging the axis weights in the matrix $\underset{\sim}{A}$, leaves the distance observations unchanged. For example when $n=2$, our estimate of $\underset{\sim}{\theta}_j$ will be an estimate of one of the elements of the set $\{(\theta_{1j}, \theta_{2j}), (-\theta_{1j}, \theta_{2j}), (\theta_{1j}, -\theta_{2j}), (-\theta_{1j}, -\theta_{2j}), (\theta_{2j}, \theta_{1j}), (-\theta_{2j}, \theta_{1j}), (\theta_{2j}, -\theta_{1j}), (-\theta_{2j}, \theta_{1j}), (-\theta_{2j}, -\theta_{1j})$. This set consists of all the reflections and 90° rotations of $\underset{\sim}{\theta}_j$. These ambiguities in the values of $\underset{\sim}{\theta}_j$ estimated are of no concern in the spatial analysis application for which the procedure was developed. Consequently, we will not be concerned with these ambiguities, but will proceed as if we are estimating $\underset{\sim}{\theta}_j$ rather than one of the other elements of the above set.

13. For values of k greater than one, however, using $\overline{D_j^4} - \overline{D_{p+1}^4}$ in the left hand side of equation (12) introduces a bias. Consider for example, $k=4$. Then

$$D_j^4 = \| \underset{\sim}{\theta}_j - \underset{\sim}{x}_r \|_A^2 + 4\| \underset{\sim}{\theta}_j - \underset{\sim}{x}_r \|_A^{3/2} \varepsilon_{jr}$$

$$+6 \| \underset{\sim}{\theta}_j - \underset{\sim}{x}_r \|_A \varepsilon_{jr}^2 + 4 \| \underset{\sim}{\theta}_j - x_r \|_A^{1/2} \varepsilon_{jr}^3$$

$$+\varepsilon_{jr}^4$$

Thus as long as the distribution of ε_{jr} is symmetric,

$$\overline{D_j^4} = \overline{\| \underset{\sim}{\theta}_j - x_r \|_A^2} + 6\sigma^2 \overline{\| \underset{\sim}{\theta}_j - \underset{\sim}{x}_r \|_A} + \overline{\varepsilon_{jr}^4} + O(N^{-1/2})$$

Consequently the term

$$6\sigma^2 \overline{\| \underset{\sim}{\theta}_j - x_r \|_A} + \overline{\varepsilon_{jr}^4}$$

must be subtracted from $D_j^4 - D_{p+1}^4$ for each j. To do this we assume that the second term is small relative to the first. Since

$$\overline{D_j^2} = \overline{\| \theta_j - x_r \|_A} + 2\overline{\| \theta_j - x_r \|_A^{1/2} \varepsilon_{jr}} + \overline{\varepsilon_{jr}^2}$$

$$= \overline{\| \theta_j - x_r \|_A} + \sigma^2 + O(N^{-1/2})$$

we can estimate $\| \theta_j - x_r \|_A$ by $\overline{D_j^2} - \hat{\sigma}^2$ where $\hat{\sigma}$ is obtained from Result 2, and, in turn, estimate $6\sigma^2 \| \theta_j - x_r \|_A$ by $6\hat{\sigma}^2 (\overline{D_j^2} - \hat{\sigma}^2)$. This is subtracted from $D_j^4 - D_{p+1}^4$.

14. See, for example, Kenneth Hoffman and Ray Kunze, Linear Algebra (N. J.: Prentice-Hall, 1961), p. 259.

15. Special consideration must be given to the possibilities, $\alpha_2 = 0$, $\xi = 45° + 90°i$ $(i=0,+1,...)$ and $\xi = 90° + 90°i$ $(i=0,+1,...)$ lest we attempt to divide by zero. To handle these possibilities, after performing the regression we test for $\alpha_2 = 0$ and $\alpha_1 = \alpha_2$. Using the results of this test we consider four exhaustive cases:

Case I: $\alpha_2 \neq 0$ and $\alpha_1 \neq \alpha_2$. Since $\alpha_2 \neq 0$ then $a_1 \neq a_2$ and $\xi_1 = 90i$, $i = 0, \pm1, \ldots$. Since $\alpha_1 \neq \alpha_2$, $\xi \neq 45° + 90°i$, $(i=0,\pm1 \ldots$. Thus, we encounter no problematic possibilities.

Case II: If $\alpha_2 = 0$ and $\alpha_1 \neq \alpha_2 = 0$, we have $a_1 = a_2$, or $\xi = 90°i$, $i = 0, \pm1, \ldots$, or both. But since $\alpha_1 \neq \alpha_2$, $a_1 \neq a_2$. Thus, $\xi = 90°i$, $i = 0, \pm1, \ldots$. Choose $\xi = 0$ (if $\xi = 90°, 180°, 270°$, the result is simply a corresponding rotation in the recovered positions), in which case $a_1 = \alpha_0^{-1}$, $a_2 = \alpha_1^{-1}$, $\overline{x}_1 = -\alpha_3 a_1^{1/2}/2$, and $\overline{x}_2 = -\alpha_4 a_2^{1/2}/2$.

Case III: If $\alpha \neq 0$ and $\alpha_1 = \alpha_2$, then $\xi = 45° + 90°i$, $i = 0, \pm1, \ldots$, and we have $\alpha_0 = \alpha_1 = 1/2(a_1^{-1} + a_2^{-1})$ and $\alpha_2 = a_2^{-1} - a_1^{-1}$.

Case IV: If $\alpha_2 = 0$ and $\alpha_1 = \alpha_2$, then $a_1 = a_2$ in which case we cannot identify ξ, since regardless of what value we assign to ξ our recovered positions are a simple rotation of their true positions.

16. Donald Stokes, "Spatial Models of Party Competition," American Political Science Review, 57 (June, 1963).

17. Aside from the desirability of estimating distances as opposed to relative positions, one advantage of our procedure is that the estimates of the issue weights a_i provide an internal check for the possible existence of valence issue. Rabinowitz and Rusk and Weisberg do not consider valence issues because their nonmetric method would not indicate the presence of such an issue. Here, however, a negative $\hat{a}_i$ indicates a misspecification in the model underlying the methodology. If the valence issue had not worked with a politically meaningful grouping, the spatial model would have been rejected.

18. Note that while Figure 1a conforms to a utility function that renders citizens risk averse, attitudes towards risk become less clear for $k > 1$. As we note earlier, however, several conceptual difficulties are encountered if we seek to interpret expression (2) as a utility function -- the principal difficulty being the absence of a common cardinally defined space. Here, then, the simplest solution is to assume without formal justification that on the average citizens are risk averse.

While a more rigorous treatment of risk seem warranted in the context of scaling methodologies, the data here is far too crude to admit a more refined analysis.

19. The income, educational levels and voting statistics for the subsample and the total sample of respondents are as follows:

Income	Total Sample	Subsample
Less than $1,999	9.0%	4.2%
$2,000-3,999	14.2	10.4
$4,000-5,999	13.4	11.1
$6,000-7,999	18.4	19.6
$8,000-9,999	13.0	16.3
$10,000-11,999	11.1	11.9
$12,000-14,999	9.5	12.2
$15,000-19,999	6.0	7.0
$20,000-24,999	2.0	2.6
$25,000 or more	3.3	4.6
Educational Level		
Eight Grades or less	21.3%	13.0%
Between Eight and Twelve	40.0	35.2
Some or all of college	34.6	45.3
Advanced Degrees	4.1	6.5
Voting		
Voted	75.8%	86.5%
Abstained	24.2%	13.5%

20. The sample is actually divided into three overlapping subsets. "Democrats" are those who identified themselves as weak and strong Democrats and Independents leaning Democratic. "Independents" are those who identified themselves as Independents and Independents leaning Republican or Democratic. The definition of "Republican" parallels that for Democrats.

21. Recall that $\underset{\sim}{W}$ is chosen such that the mean of the last $p - m$ eigenvalue is zero. The dimensionality of the space is m. In this case, some of the eigenvalues are negative. The percentages are based on the sum of the eigenvalues; thus, the numbers sum to more than 100 percent in some cases.

22. Alternatively, we can extract two dimensions from the factor analysis and order the candidates over the issue along which their variations are most substantial. This changes the order somewhat since, by extracting only a single dimension, we cannot correct for rotational ambiguities. Briefly, the order we obtain using this second procedure for Republicans and Democrats is

Republicans	Democrats
Johnson	Johnson
Humphrey	Humphrey
Muskie	Muskie
Kennedy	Kennedy
Romney	Rockefeller
Rockefeller	Nixon
Nixon	McCarthy
McCarthy	Romney
Agnew	Agnew
Reagan	Reagan

These orderings differ somewhat from those we present in Table 3 (especially among Republicans and the placement of Nixon by Democrats). Nevertheless, roughly the same pattern emerges.

23. While a one dimensional configuration of candidates differ substantially from the two-dimension recovered by Rabinowitz and by Rusk and Weisberg, it parallels the conclusions of Aldrich and McKelvey. Briefly, using perceptual rather than preference data on several substantive (actionable) issues, they find that the ordering of the candidates is nearly identical on every issue -- thereby suggesting an underlying unidimensional partisian or left-right continuum. Rusk and Weisberg, op. cit.; Rabinowitz, op. cit.; John H. Aldrich and Richard D. McKelvey, "A Method of Scaling, with Application to the 1968 and 1972 Presidential Elections,: paper presented at APSA annual meetings, Chicago, 1974.

24. Personal communication.

25. In these and all subsequent figures, distances are scaled by the estimated a's. That is, the relative saliencies of the dimensions should be interpreted as unity, i.e., $a_1/a_2 = 1$.

26. The data used here is the post election response to how a citizen voted or, if he abstained, how he would have voted if he voted.

APPENDIX

The statement of Result 1 depends on k. In the text we set k = 1, but here we derive the result for k = 4 which we used in the survey analysis. Since

$$D_{jr} = \| \underset{\sim}{\theta}_j - \underset{\sim}{x}_r \|_{\underset{\sim}{A}}^{1/2} + \varepsilon_{jr},$$

we have

$$R_{jr} = D_{jr}^4 = (\underset{\sim}{\theta}_j - \underset{\sim}{x}_r)\underset{\sim}{A}(\underset{\sim}{\theta}_j - \underset{\sim}{x}_r) + \delta_{jr}, \tag{A1}$$

where

$$\delta_{jr} = 4\| \underset{\sim}{\theta}_j - \underset{\sim}{x}_r \|_{\underset{\sim}{A}}^{3/2}\varepsilon_{jr} + 6\| \underset{\sim}{\theta}_j - \underset{\sim}{x}_r \|_{\underset{\sim}{A}}\varepsilon_{jr}^2 + 4\| \underset{\sim}{\theta}_j - \underset{\sim}{x}_r \|_{\underset{\sim}{A}}^{1/2}\varepsilon_{jr}^3 + \varepsilon_{jr}^4. \tag{A2}$$

We begin the construction of the covariance matrix from the distances

$$Y_{jr} = R_{jr} - R_{p+1,r} = \| \underset{\sim}{\theta}_j - \underset{\sim}{x}_r \|_{\underset{\sim}{A}}^2 - \| \underset{\sim}{\theta}_{p+1} - \underset{\sim}{x}_r \|_{\underset{\sim}{A}}^2 + \delta_{jr} - \delta_{p+1,r}. \tag{A3}$$

We desire to construct the covariance matrix of the vector $(Y_{jr},\ldots,Y_{pr})$. Thus we require the Var (Y_{jr}) and the Cov (Y_{jr},Y_{hr}). We know that the covariance matrix of $\| \underset{\sim}{\theta}_j - \underset{\sim}{x}_r \|_{\underset{\sim}{A}}^2 - \| \underset{\sim}{\theta}_{p+1} - \underset{\sim}{x}_r \|_{\underset{\sim}{A}}^2$ is $4\underset{\sim}{\theta}'\underset{\sim}{A}\underset{\sim}{\Sigma}\underset{\sim}{A}\underset{\sim}{\theta}$.

Thus we have

$$\begin{aligned}\text{Var } Y_{jr} = {} & 4\underset{\sim}{\theta}_j'\underset{\sim}{A}\underset{\sim}{\Sigma}\underset{\sim}{A}\underset{\sim}{\theta}_j + \text{Var }(\delta_{jr}) + \text{Var }(\delta_{p+1,r}) - 2\text{ Cov }(\delta_{jr},\delta_{p+1,r}) \\ & + 2\text{ Cov }(\delta_{jr},\| \underset{\sim}{\theta}_j - \underset{\sim}{x}_r \|_{\underset{\sim}{A}}^2) - 2\text{ Cov }(\delta_{jr},\| \underset{\sim}{\theta}_{p+1} - \underset{\sim}{x}_r \|_{\underset{\sim}{A}}^2) \\ & - 2\text{ Cov }(\delta_{p+1,r},\| \underset{\sim}{\theta}_j - \underset{\sim}{x}_r \|_{\underset{\sim}{A}}^2) + 2\text{ Cov }(\delta_{p+1,},\| \underset{\sim}{\theta}_{p+1} - \underset{\sim}{x}_r \|_{\underset{\sim}{A}}^2)\end{aligned} \tag{A4}$$

and

$$\begin{aligned}\text{Cov }(Y_{jr},Y_{kr}) = {} & 4\underset{\sim}{\theta}_j'\underset{\sim}{A}\underset{\sim}{\Sigma}\underset{\sim}{A}\underset{\sim}{\theta}_k + \text{Cov }(\delta_{jr},\| \underset{\sim}{\theta}_k - \underset{\sim}{x}_r \|_{\underset{\sim}{A}}^2) \\ & - \text{Cov }(\delta_{jr},\| \underset{\sim}{\theta}_{p+1} - \underset{\sim}{x}_r \|_{\underset{\sim}{A}}^2) - \text{Cov }(\delta_{p+1,r},\| \underset{\sim}{\theta}_k - \underset{\sim}{x}_r \|_{\underset{\sim}{A}}^2) \\ & + \text{Cov }(\delta_{p+1,r},\| \underset{\sim}{\theta}_{p+1} - \underset{\sim}{x}_r \|_{\underset{\sim}{A}}^2) + \text{Cov }(\delta_{kr},\| \underset{\sim}{\theta}_j - \underset{\sim}{x}_r \|_{\underset{\sim}{A}}^2) \\ & - \text{Cov }\delta_{kr},(\| \underset{\sim}{\theta}_{p+1} - \underset{\sim}{x}_r \|_{\underset{\sim}{A}}^2) - \text{Cov }(\delta_{p+1,r},\| \underset{\sim}{\theta}_j - \underset{\sim}{x}_r \|_{\underset{\sim}{A}}^2) \\ & + \text{Cov }(\delta_{p+1,r},\| \underset{\sim}{\theta}_{p+1} - \underset{\sim}{x}_r \|_{\underset{\sim}{A}}^2) + \text{Cov }(\delta_{jr},\delta_{kr}) \\ & - \text{Cov }(\delta_{jr},\delta_{p+1,r}) - \text{Cov }(\delta_{kr},\delta_{p+1,r}) \\ & + \text{Var }(\delta_{p+1,r}).\end{aligned} \tag{A5}$$

The covariance matrix of $\underset{\sim}{Y}_{rj}$ is

$$4\underset{\sim}{\theta}'\underset{\sim}{A}\underset{\sim}{\Sigma}\underset{\sim}{A}\underset{\sim}{\theta} + \underset{\sim}{D},$$

where $\underset{\sim}{D} = (d_{jk})$ and d_{jk} is defined by expressions (A4) and (A5).

What remains, then, is a proof of Result 2 -- our procedure for estimating σ^2.

We give here the proof of Result 2 for k = 1, since for larger k we encounter considerable notational complexity. The basic theme of the argument applies, however, for k > 1. Letting $\sigma^2 = \hat{\sigma}^2 + \delta$, then, for k = 1, we obtain from Result 1 and the definition of $\hat{W}$,

$$\hat{W} = \theta' A \Sigma A \theta + \delta I + \delta 11' + \varepsilon \qquad \text{(A6)}$$

Consider now the cases of $\delta \geq 0$ and $\delta < 0$ separately.

(i) $\delta \geq 0$. Since, by construction, the smallest eigenvalue of $\hat{W}$ is zero, we have for $x'x = 1$,

$$\min_x x'\hat{W}x = \min_x [x'\theta A \Sigma A \theta x + \varepsilon x' I x + \varepsilon x' 11' x + x'\varepsilon x] = 0 \qquad \text{(A7)}$$

Since $A\Sigma A$ and $11'$ are singular positive semi-definite matices, $\min x'11'x = \min x'\theta A \Sigma A \theta x = 0$ and $\min x'\varepsilon x = 0(N^{-1/2})$, we have

$$\min_x x'\hat{W}x \geq \varepsilon + 0(N^{-1/2}) \geq 0$$

But, since $\min x'\hat{W}x = 0$, δ must then be of order $0(N^{-1/2})$.

(ii) $\delta < 0$. Returning to expression (A7), we know that there exists an x such that $x'\theta' A \Sigma A \theta x = 0$ and that $x'Ix = 1$, $x'11'x \geq 0$ and $x'\varepsilon x = 0(N^{-1/2})$. Thus, if $\delta < 0$, we have

$$\min x'\hat{W}x \leq \delta + 0(N^{-1/2}) \leq 0$$

But, again since $\min x'\hat{W}x = 0$, ε must be of order $0(N^{-1/2})$.

This establishes, then, that if we choose $\hat{\sigma}$ such that the smallest eigenvalue of $\hat{W}$ equals zero,

$$\hat{\sigma}^2 = \sigma^2 + 0(N^{-1/2}),$$

or equivalently that $\hat{W} = \theta' A \Sigma A \theta + \varepsilon$ where ε is a symmetric matrix of order $0(N^{-1/2})$.[25]