Bayesian Metric Multidimensional Scaling

Ryan Bakker

*School of Public and International Affairs, Department of Political Science,
University of Georgia, Athens, GA*
e-mail: rbakker@uga.edu (corresponding author)

Keith T. Poole

*School of Public and International Affairs, Department of Political Science,
University of Georgia, Athens, GA*
e-mail: ktpoole@uga.edu

Edited by R. Michael Alvarez

In this article, we show how to apply Bayesian methods to noisy ratio scale distances for both the classical similarities problem as well as the unfolding problem. Bayesian methods produce essentially the same point estimates as the classical methods, but are superior in that they provide more accurate measures of uncertainty in the data. Identification is nontrivial for this class of problems because a configuration of points that reproduces the distances is identified only up to a choice of origin, angles of rotation, and sign flips on the dimensions. We prove that fixing the origin and rotation is sufficient to identify a configuration in the sense that the corresponding maxima/minima are inflection points with full-rank Hessians. However, an unavoidable result is multiple posterior distributions that are mirror images of one another. This poses a problem for Markov chain Monte Carlo (MCMC) methods. The approach we take is to find the optimal solution using standard optimizers. The configuration of points from the optimizers is then used to isolate a single Bayesian posterior that can then be easily analyzed with standard MCMC methods.

1 Introduction

In this article, we take a fresh look at the classical ratio scale similarities and unfolding problems from the psychometrics literature using Bayesian methods. Similarities and unfolding are multidimensional scaling (MDS) methods. MDS encompasses a wide variety of statistical techniques aimed at characterizing structure within a set of preference or perceptual data. The most common uses of MDS are to uncover the dimensionality of a given set of data and to visually display the placements of stimuli (i.e., products, candidates, etc.) according to their positions on the dimension(s) (this is the similarities problem). Additionally, there is often interest in the placement of respondents who have expressed preferences for the stimuli and therefore can be located on the dimension(s) relative to these stimuli (this is the unfolding problem). Such spatial maps can help us better understand the structure of certain types of decision-making metrics employed in a variety of settings. This information can then be used to predict the outcome of future choices (e.g., what product a certain consumer will buy or what candidate a voter will support), as an individual will be more likely to choose a stimulus that is closer to her position on a given dimension than one that is farther away.

In political science, MDS has a natural connection to spatial (geometric) models of choice and judgment. In the basic spatial model, individuals/decision-makers are assumed to have single-peaked utility functions over some latent dimensions of judgment and they choose the stimuli closest to them on the dimension(s) (Enelow and Hinich 1984). For example, the methods developed in the past 30 years to estimate ideological positions from roll call data are examples

---

Authors' note: Supplementary materials for this article are available on the *Political Analysis* Web site.

of multidimensional unfolding applied to explicitly stated spatial models of choice (e.g., Poole and Rosenthal 1997; Londregan 2000; Martin and Quinn 2002; Clinton, Jackman, and Rivers 2004; Pope and Treier 2011).

In contrast to these modern methods that analyze nominal (yea or nay) data, the first uses of MDS methods in political science were to ratio scale data. For example, Weisberg and Rusk (1970) used nonmetric MDS (Shepard 1962a, 1962b; Kruskal 1964a, 1964b) to analyze a correlation matrix computed over respondents' placements of political stimuli on 100-point "feeling" thermometers. Later Rabinowitz (1976), Cahoon, Hinich, and Ordeshook (1978), Jacoby (1982), and Brady (1990) developed unfolding methods based on the spatial theory of choice to analyze feeling thermometers.

We revisit these problems and show how to apply Bayesian methods to noisy ratio scale distances for both the similarities and unfolding problems. Our approach is explicitly based on the spatial theory of choice and judgment. Since these problems have been studied for ≥ 50 years, the solutions are known and various data sets have been used to calibrate a succession of statistical methods. Both similarities and unfolding can easily be handled with frequentist or Bayesian models, but pose problems for Markov chain Monte Carlo (MCMC) methods because of the existence of multiple posterior distributions that are mirror images of one another over the hyperplane of the parameters. In the Supplementary Appendix, we prove two theorems and two corollaries concerning the identification of posterior distributions in distance-based choice models that allow us to isolate a single posterior distribution.

The analysis of ratio scale similarities data by psychometricians in the 1930s through the 1960s led to the development of MDS methods. The psychometricians solved the general problem of representing relational or distance data in a spatial or geometric map where the points represented the stimuli and the distances between the points in the geometric map reproduced the observed distance/relational data. The ratio scale similarities problem was solved by Torgerson (1952, 1958), which in turn built upon work done by pschometricians in the 1930s (Eckart and Young 1936; Young and Householder 1938).

In the unfolding problem, there are two sets of points—one representing individuals and the other representing stimuli. The observed distance/relational data are regarded as expressing the preferences of individuals; namely, the closer a stimulus point is to an individual point, the more the individual prefers that stimulus. The unfolding problem for ratio scale data (the "metric unfolding problem") was first solved by Schönemann (1970).

We first discuss the similarities problem, and then we turn to the unfolding problem. Since our Bayesian framework is essentially the same for both problems, we spend more time detailing our solution for the similarities problem because the mathematical exposition is simpler. However, the unfolding problem is of greater interest because most public opinion survey data sets include a set of relational data questions in some form ("where would you place George Bush?", "On a scale of zero to ten, how would you rate John Kerry?", etc.).

2   Multidimensional Bayesian Similarities/Dissimilarites Scaling

Similarities data differ from choice data in that they measure how alike or not-alike objects are to each other. Similarities data are frequently generated from choice data (e.g., an agreement score matrix based on how often legislators vote together), but do not include the choices themselves.

These data are relational and organized as a square matrix. For example, imagine you have never seen a map of the United States, but you know the distances in miles between a set of U.S. cities. Within these distances is embedded a map, and the purpose of MDS methods is producing a map given the distances; that is, in this instance, placing a set of points representing the cities in a plane such that they reproduce the distances. In this case, MDS produces a geographical map that represents the distances in a meaningful way (i.e., the map would be useful for getting from one city to another). Clearly, a two-dimensional map of distances between cities would be a simplification of the true positions of the cities as we exist in a three-dimensional world, but it is good enough for driving from one city to the next.

Table 1 Driving distances between five cities

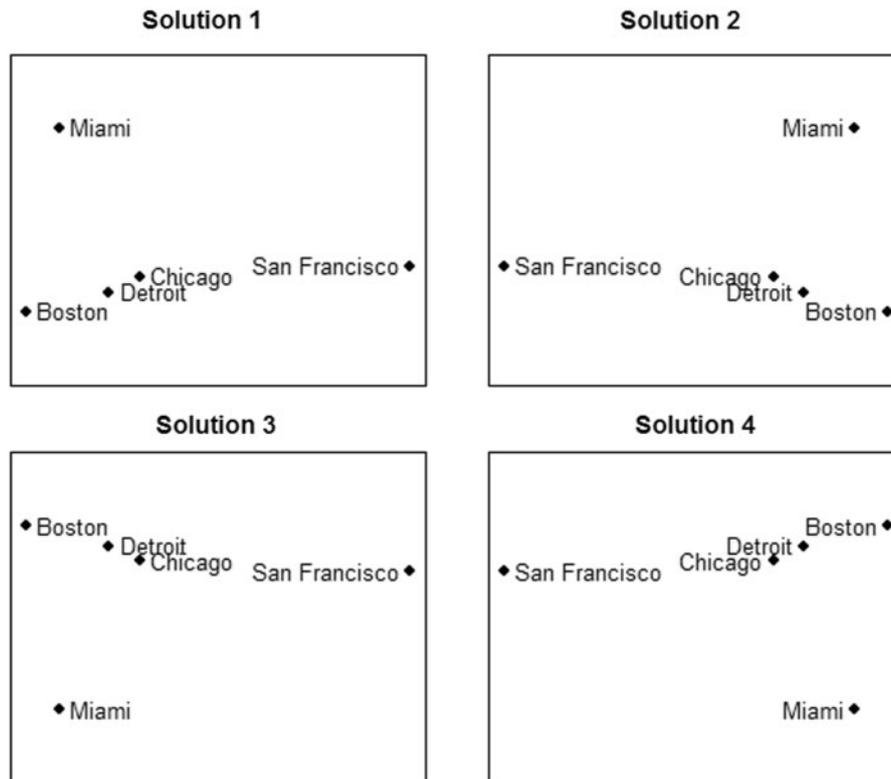| City	Boston	Detroit	Chicago	SF	Miami
Boston	0	702	983	3179	1539
Detroit	702	0	279	2475	1409
Chicago	983	279	0	2212	1309
SF	3179	2475	2212	0	3097
Miami	1539	1409	1390	3097	0

Fig. 1 Solutions for five cities' data.

Table 1 displays the distances in miles between five major U.S. cities. With just these data, we can generate a map using MDS techniques. However, there is an additional complication inherent in MDS techniques. The cities in Table 1 can be arranged in any number of different configurations that accurately represent the distances between them. Indeed, there are an infinite number of arrangements of the points that reproduce the distances. As we describe below, this issue requires that we impose some constraints in order to produce a single, meaningful graphical representation, and it also has additional implications for the Bayesian approach we present. In Fig. 1, we plot four solutions from an MDS routine applied to the city distance data.

To anyone with cursory knowledge of U.S. geography, solution 4 is obviously the true configuration. Clearly, there is no need to estimate the positions of the U.S. cities based on their distances, as we know the exact positions of their coordinates (i.e., we know their latitude and longitude). Most applications of MDS to social science data, however, involve estimating positions on dimensions that are subjective in nature, for example, left–right ideological positions of the legislators. As there is no "true" set of coordinates in such settings, the relative placement of the points (as illustrated in Fig. 1) will always be arbitrary. In addition, the observed data will be

noisy. Consequently, given a relative placement, it is highly desirable that the uncertainty in the distances is reflected in the estimated coordinates. In other words, we are interested not only in the relative positions of the certain objects (i.e., legislators) but also in how certain we are about these estimated positions. This is particularly the case when using MDS solutions for statistical inference.

In psychology, various methods of MDS have been developed during the past 60 years to analyze the similarities data. MDS methods model these similarities as distances between points in a geometric space (usually simple Euclidean). These MDS programs were designed to produce a picture or spatial map that summarizes the data graphically. MDS analyses are almost always in three dimensions or less because the whole *raison d'être* of MDS methods is to produce a visual summary. In this sense, a matrix of similarities/dissimilarities data can be regarded as having a geometric map embedded within it.

The purpose of an MDS method is to recover this geometric map. This requires a set of assumptions about the dissimilarities that allow them to be represented as points in a Euclidean space. The simplest assumption is just to model the observed dissimilarities as noisy realizations of the underlying true distances. For example, denote the observed dissimilarity (distance) as d_{jm}^*:

$$d_{jm}^* = d_{jm} + \varepsilon_{jm}, \tag{1}$$

where j and m are both indices for the stimuli; i.e., $j = 1, \ldots, q$; $m = 1, \ldots, q$. Let Z_{jk} be the jth stimulus coordinate on the kth dimension, $k = 1, \ldots, s$, where s is the number of dimensions. Let d_{jm} be the Euclidean distance between stimulus j and stimulus m in the s-dimensional space:

$$d_{jm} = \sqrt{\sum_{k=1}^{s} (Z_{jk} - Z_{mk})^2} \tag{2}$$

and

$$\varepsilon_{jm} = d_{jm}^* - d_{jm} \sim N(0, \sigma^2) \tag{3}$$

or

$$\varepsilon_{jm} \sim \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{-\frac{1}{2\sigma^2}\left(d_{jm}^* - \sqrt{\sum_{k=1}^{s}(Z_{jk} - Z_{mk})}\right)^2}.$$

This produces the likelihood function:

$$\mathbf{L}^*(\mathbf{Z}_{jk}|\mathbf{D}^*) = \frac{1}{(2\pi\sigma^2)^{q(q-1)/2/2}} e^{-1/2\sigma^2 \sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(d_{jm}^* - \sqrt{\sum_{k=1}^{s}(Z_{jk} - Z_{mk})^2}\right)^2}, \tag{4}$$

where $\boldsymbol{D}^*$ is the $q \times q$ matrix of observed dissimilarities. Taking the log of the right-hand side and dropping the unnecessary constants yields a standard squared error loss function:

$$\ell n\xi = -\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(d_{jm}^* - \sqrt{\sum_{k=1}^{s}(Z_{jk} - Z_{mk})^2}\right)^2 = -\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(d_{jm}^* - d_{jm}\right)^2. \tag{5}$$

In standard optimization methods, the first derivatives of equation (5) are used to find maximum likelihood estimates of the points. However, these derivatives are problematic because they all contain the ratio d_{jm}^*/d_{jm}, which is undefined when $Z_j = Z_m$ so that $d_{jm} = 0$. In practice, this is not a problem, but it and the fact that distances cannot be negative means that the statistical properties are not clear and that the assumption about the error, equation (3), is dicey at best. Nevertheless, finding Z's that minimize (or maximize as in equation (5)) the squared error loss function is relatively easy.

We now turn to a more realistic model of the data. We assume that the observed distances, d^*_{jm}, are drawn from the log-normal distribution because distances are inherently positive:

$$\ln(d^*_{jm}) \sim N(\ln(d_{jm}),\sigma^2). \tag{6}$$

That is,

$$f(d^*_{jm}) = \frac{1}{(2\pi\sigma^2)^{1/2}d^*_{jm}} e^{\left(-1/2\sigma^2 \left(\ln(d^*_{jm})-\ln(d_{jm})\right)^2\right)}.$$

Other researchers have used the truncated normal (Oh and Raftery 2001), the normal (Navarro and Lee 2003), and the normal with an exponential mean (Okada and Shigemasu 2010). We prefer the log-normal because we think it is a more realistic model of the noise process; namely, the smaller the observed distance, the smaller the variance of that distance.

Our likelihood function is

$$\mathbf{L}^*(Z_{jk}|\mathbf{D}^*) = \frac{1}{(2\pi\sigma^2)^{q(q-1)/2/2}} \left(\prod_{j=1}^{q-1} \prod_{m=j+1}^{q} \frac{1}{d^*_{jm}}\right) e^{-1/2\sigma^2 \sum_{j=1}^{q-1} \sum_{m=j+1}^{q} \left(\ln\left(d^*_{jm}\right)-\ln\left(\sqrt{\sum_{k=1}^{s}(Z_{jk}-Z_{mk})^2}\right)\right)^2}. \tag{7}$$

To implement our Bayesian model, we use simple normal prior distributions for the stimuli coordinates:

$$\xi(Z_{jk}) = \frac{1}{(2\pi\kappa^2)^{1/2}} e^{-Z^2_{jk}/2\kappa^2} \tag{8}$$

and a uniform prior for the variance term:

$$\xi(\sigma^2) = \frac{1}{c}, \ 0 <c< b, \tag{9}$$

where, empirically, b is no greater than 2.

Hence, our posterior distribution is

$$\xi(Z_{jk}|\mathbf{D}^*) \propto \prod_{j=1}^{q-1} \prod_{m=j+1}^{q} \left\{f_{jm}(Z_{jm}|d^*_{jm})\right\}\xi(Z_{11})\xi(Z_{12})...\xi(Z_{1s})\xi(Z_{21})...\xi(Z_{qs})\xi(\sigma^2). \tag{10}$$

Taking the log of the right-hand side and dropping the unnecessary constants:

$$\ln\xi \propto -\frac{q(q-1)/2}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=1}^{q-1}\sum_{m=j+1}^{q} \left(\ln\left(d^*_{jm}\right) - \ln\left(\sqrt{\sum_{k=1}^{s}(Z_{jk}-Z_{mk})^2}\right)\right)^2$$
$$-\frac{1}{2\kappa^2}\left(\sum_{j=1}^{q}\sum_{k=1}^{s}Z^2_{jk}\right) - \ln(c)= \tag{11}$$
$$-\frac{q(q-1)/2}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=1}^{q-1}\sum_{m=j+1}^{q}\left(\ln\left(d^*_{jm}\right) - \ln\left(d_{jm}\right)\right)^2 - \frac{1}{2\kappa^2}\left(\sum_{j=1}^{q}\sum_{k=1}^{s}Z^2_{jk}\right) - \ln(c).$$

We experimented with vague and informative priors and found that the solutions were essentially identical for large matrices. We discuss this in more detail below. In the Supplementary Appendix, we show the first and the second derivatives for equation (11). In our estimation work, we check the solutions with both numerical and analytical first and second derivatives.

To illustrate our approach to similarities scaling, we use agreement scores computed between members of the US 90th (1967–68) Senate. We chose the 90th Senate because it is well known that voting was two-dimensional during this period (Poole and Rosenthal 1997). Given q roll call votes, the agreement score is the number of times a pair of senators vote the same way (yea, yea or nay, nay) divided by the number of times that they both voted on the same roll calls and multiplied by

Table 2 Agreement scores for 90th Senate (partial)

JOHNSON (D-Pres)	100	61	50	52	65	70	37	–
SPARKMAN (D-AL)	61	100	89	50	65	85	65	–
HILL (D-AL)	50	89	100	53	62	78	69	–
GRUENING (D-AK)	52	50	53	100	76	58	43	–
BARTLETT (D-AK)	65	65	62	76	100	70	47	–
HAYDEN (D-AZ)	70	85	78	58	70	100	57	–
FANNIN (R-AZ)	37	65	69	43	47	57	100	–

100. The agreement scores range from 0 to 100, with 100 indicating identical voting records. Table 2 shows a few senators and their agreement scores.

We convert the agreement scores to distances by subtracting them from 100 and dividing by 50. This is a convenient normalization because the estimated coordinates are usually in the unit hypersphere. Note that we include President Lyndon Johnson in the matrix by using *Congressional Quarterly's* presidential support roll calls. That is, CQ indicates on a fair number of roll calls whether a yea/nay is a vote in favor of the president's position. Hence, the president can be treated as a senator. He just does not vote as often.

In this example, counting President Johnson, q is equal to 102. Our $q \times q$ symmetric matrix of distances has $q(q-1)/2$ unique entries (we ignore the diagonal of zeros). Suppose there is an exact solution; that is, a set of q points in s dimensions that exactly reproduces the distances. Clearly, given that we only observe the distances, it does not matter what origin, rotation around that origin, or sign flips on the dimensions we select as long as the configuration of points vis-à-vis one another is not altered.

With q points in s dimensions, we have to solve for $q * s$ coordinates. However, we can set any point to the origin—$(0, 0, \ldots, 0)$—so this leaves us with $q * s - s = (q - 1) * s$ parameters. To pin down the configuration, we need to set the rotation. In general, in s dimensions, with one point set to the origin, a rigid rotation of the configuration is determined by $s - 1$ angles from the origin. In addition, given $s - 1$ fixed angles in the rotation matrix, if we have a solution that exactly reproduces the matrix of distances, then there are an additional $2^s - 1$ solutions that exactly reproduce the distances. These additional solutions are simply sign flips on the dimensions. For example, with $s = 2$, suppose that we have a solution such that it reproduces our matrix of distances. Then there are three more solutions that also exactly reproduce the distances.

To fix the origin, we set Senator Hill (D-AL) at the origin and we fix President Johnson's second dimension coordinate at zero, which has the effect of fixing the angle from the origin in the rotation matrix. We use the Nelder-Mead (1965) amoeba method and the Powell (1973) method to obtain 1001 solutions for our log-posterior, equation (11), from random starts. The best solution and its reflections are shown in Fig. 2. The tokens in the plots indicate the political party of the member—"D" for northern Democrat, "S" for southern Democrat, and "R" for Republican. We computed both numerical and analytical first and second derivatives for the optimal solution to check that the Hessian was full rank (i.e., negative definite; see Supplementary Appendix).

Of the 1001 solutions for the Bayes posterior, only three were the solution (and its reflections) shown in Fig. 2. The log-likelihood was about −3100.0. The value for σ^2 was 0.1104. The extreme nonlinearity of the log-normal likelihood function meant that a large number of modes were found by the optimizers. Many of these were quite close together in terms of log-likelihood.

In the two dimensions, fixing the origin and one coordinate of another point at zero is enough to pin down four identical posteriors corresponding to the sign flips, as shown in Fig. 2. This is enough so that an optimizer can find modes but not enough for the efficient use of MCMC methods because of the reflections. That is, each member of the chain is a configuration. If the chain explores the entire hyperplane of the parameters, then the means of all the coordinates will be zero because of the symmetry of the modes. However, in practice, we can let the chain wander
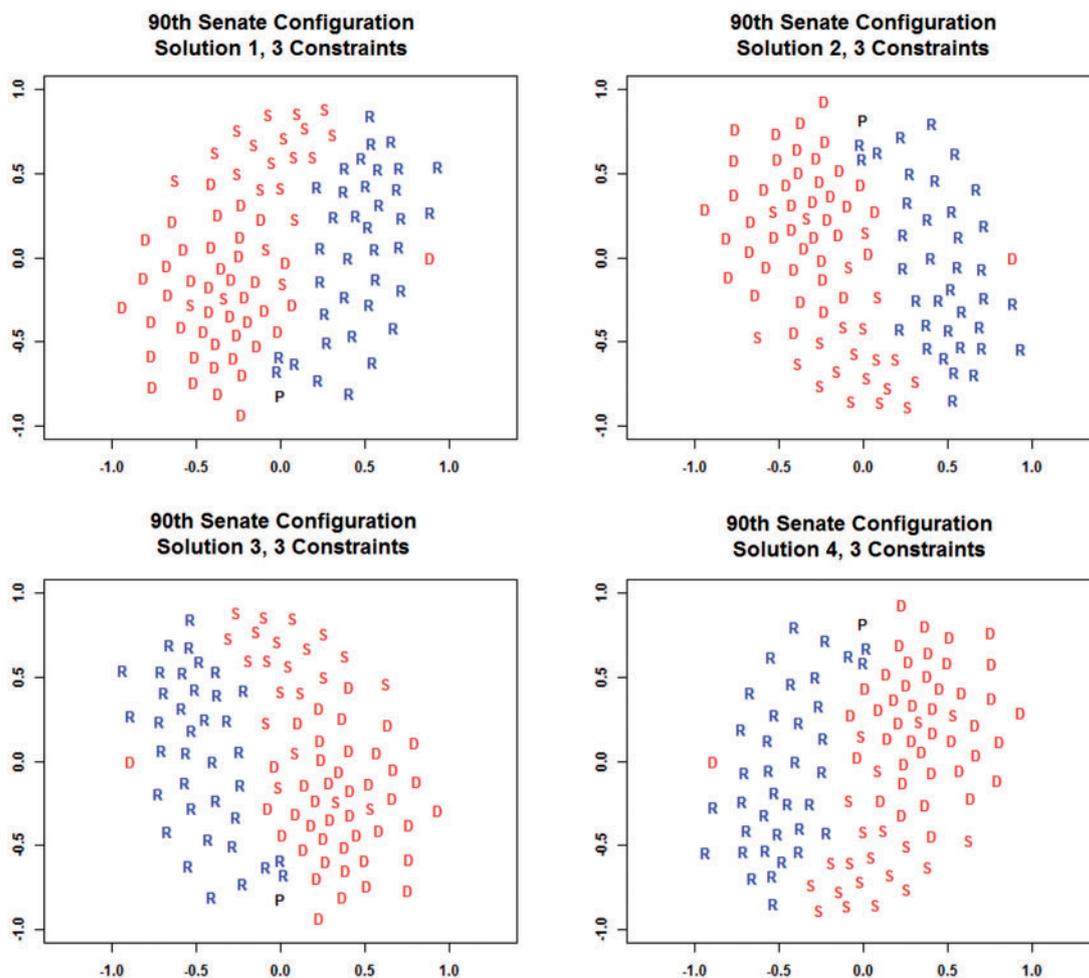
Fig. 2 Best 90th Senate configuration and its reflections.

through the $((q-1)*s)-1$ dimensional hyperplane and postprocess the results by flipping the signs of each configuration in the chain back to a target configuration. This approach is very similar to that advocated by Bradlow and Schmittlein (2000), Oh and Raftery (2001), Hoff, Raftery, and Handcock (2002), and Gormley and Murphy (2006).

For small similarities problems, we found that in addition to the origin and one fixed coordinate, simply adding three sign constraints to the three fixed coordinates isolated a single posterior. That is, keep the three constraints used to find the modes and then restrict three other coordinates to be positive/negative. This works well, and it is easy to implement in WinBUGS by using the I(,0) or I(0,) operators.

For larger problems like the 90th Senate agreement scores, we retain the origin and one fixed coordinate and then solve for the sign flips by computing simple correlations dimension by dimension between coordinates from each draw in the chain (a configuration of points), and the coordinates from the optimizer solution.

The left panel of Fig. 3 shows the results for the 90th Senate (we adjusted the coordinates to -1 to $+1$ for presentation purposes).[1] We ran multiple chains using a slice sampler (Neal 2003) for 110,000 iterations and treated the first 10,000 draws as burn-in. The configuration is the mean of draws 10,001–110,000 (all the chains produced essentially identical results). The configuration is similar to that shown in Fig. 2. The variance term is precisely estimated with a standard deviation

---

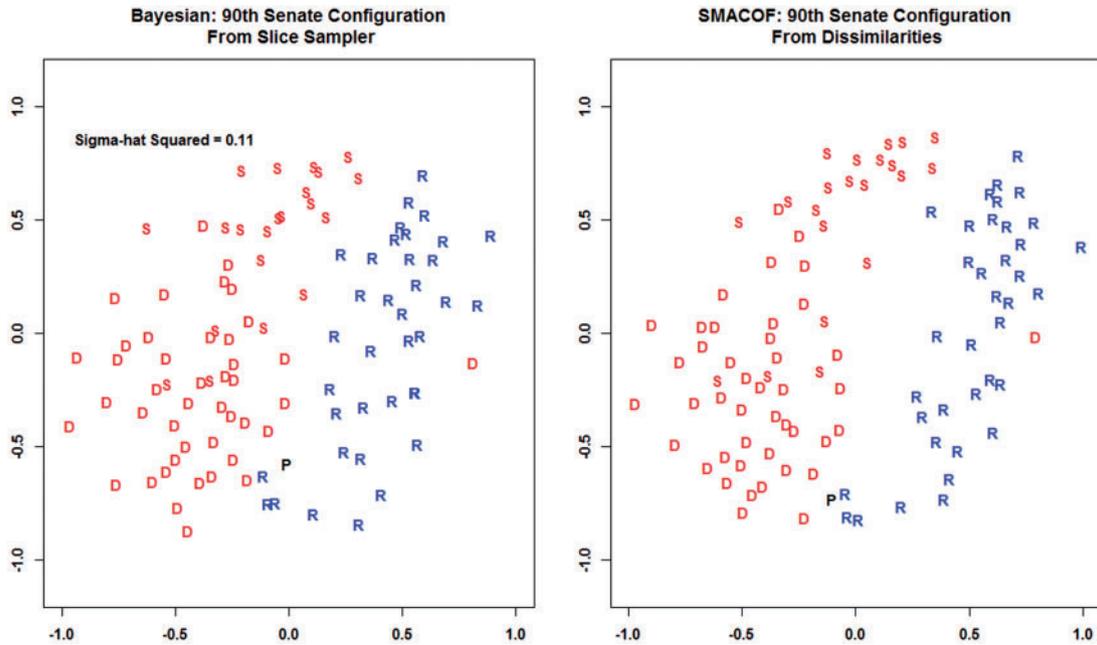[1] See Bakker and Poole (2012) for a link to all replication materials.

Fig. 3 90th Senate Bayesian and SMACOF solutions.

(SD) of 0.0026. The SDs around the points range from about 0.08 to 0.18, with the largest being 0.25. Additionally, we assessed convergence using the Geweke, Heidelberger-Welch, and Raftery and Lewis diagnostics. According to these diagnostics, the posteriors for all parameters meet all criteria for convergence. Also, we experimented with informative priors ($\kappa^2 = 1$) and vague priors ($\kappa^2 = 100$) and the results did not change.

The right panel of Fig. 3 compares our Bayesian results with the best available MDS procedure: SMACOF (scaling by maximizing a convex function), developed by Jan De Leeuw (1977) and his colleagues (De Leeuw 1988; De Leeuw and Heiser 1988; De Leeuw and Mair 2009). SMACOF is a sophisticated scaling method, which minimizes the sum of squared error as given in equation (5) above (technically, the negative of equation (5)). SMACOF can perform nonmetric MDS (Shepard 1962a, 1962b; Kruskal 1964a, 1964b) and has a flexible structure so that missing data are easily handled (see the R implementation, De Leeuw and Mair [2009]). We also use it below in our unfolding examples.

Our Bayesian configuration is essentially identical to the SMACOF configuration despite the very different models of the data. The advantage of a Bayesian approach is that we are able to get measures of uncertainty for the points. SMACOF is a classical scaling procedure, which does not produce standard errors (SEs) for the points. However, note that in the Bayesian model fixing three coordinates has the effect of "transmitting" the uncertainty associated with those coordinates to other points. There is no solution for this. It is just inherent in the problem.

Our approach has the advantage of isolating one posterior distribution and then analyzing it with standard MCMC methods. Our approach is computationally simple and can be implemented in publicly available software, such as WinBUGS and JAGS. In the Supplementary Appendix, we show our WinBUGS script for the 90th Senate. We used informed priors derived from the Nelder-Mead configuration to stabilize the sampler in WinBUGS.

We now turn to a discussion of how to apply our approach to the unfolding problem.

3 A Bayesian Multidimensional Unfolding Model

In the unfolding problem, we have two sets of points—one for individuals and the other for stimuli. We are given only the noisy ratio scale distances between the two sets and not the distances within each set. We focus here on a Bayesian unfolding model based on the spatial model of choice. We

assume that the individuals have symmetric single-peaked utility functions. Hence, we use an *ideal point* model in contrast to a *vector* model of choice. In a vector model, the underlying utility of an individual is not single-peaked. Rather, the utility increases through the space of the stimuli along the individual's vector (direction) through the space similar to the one-dimensional IRT model. Much work has been done on Bayesian vector and mixed models of unfolding in the marketing literature (Park, DeSarbo, and Liechty 2008; Fong et al. 2010; DeSarbo, Park, and Rao 2011).

Denote the observed distance as d_{ij}^*, where

$$d_{ij}^* = d_{ij} + \varepsilon_{ij}, \tag{12}$$

where n is the number of individuals, $i = 1, \ldots, n$, and X_{ik} is the ith individual coordinate on the kth dimension. As before, let Z_{jk} be the jth stimulus coordinate on the kth dimension, $k = 1, \ldots, s$, where s is the number of dimensions. Let d_{ij} be the Euclidean distance between individual i and stimulus j in the s-dimensional space:

$$d_{ij} = \sqrt{\sum_{k=1}^{s} (X_{ik} - Z_{jk})^2}. \tag{13}$$

As before, we assume that the observed distances, d_{ij}^*, are drawn from the log-normal distribution:

$$\ln(d_{ij}^*) \sim N(\ln(d_{ij}), \sigma^2),$$

which produces the likelihood function:

$$\mathbf{L}^*(X_{ik}, Z_{jk} | \mathbf{D}^*) = \frac{1}{(2\pi\sigma^2)^{\frac{nq}{2}}} \left(\prod_{i=1}^{n} \prod_{j=1}^{q} \frac{1}{d_{ij}^*} \right) e^{\frac{-1}{2\sigma^2} \sum_{i=1}^{n} \sum_{j=1}^{q} \left(\ln\left(d_{ij}^*\right) - \ln\left(\sqrt{\sum_{k=1}^{s} (X_{ik} - Z_{jk})^2} \right) \right)^2}. \tag{14}$$

We use simple normal prior distributions for the individual and stimuli coordinates:

$$\xi(X_{ik}) = \frac{1}{(2\pi\zeta^2)^{\frac{1}{2}}} e^{\frac{-X_{ik}^2}{2\zeta^2}}$$

$$\xi(Z_{jk}) = \frac{1}{(2\pi\kappa^2)^{\frac{1}{2}}} e^{\frac{-Z_{jk}^2}{2\kappa^2}}$$

and an uniform prior for the variance term:

$$\xi(\sigma^2) = \frac{1}{c}, \ 0 < c < b,$$

where b, empirically, is no greater than 2.

Hence, our posterior distribution is

$$\xi(X_{ik}, Z_{jk} | \mathbf{D}^*) \propto \prod_{i=1}^{n} \prod_{j=1}^{q} \left\{ f_{ij}(X_{ik}, Z_{jk} | d_{ij}^*) \right\} \xi(X_{11}) \ldots \xi(X_{ns}) \xi(Z_{11}) \ldots \xi(Z_{qs}) \xi(\sigma^2). \tag{15}$$

Taking the log of the right-hand side and dropping the unnecessary constants,

$$\ell n \xi \propto -\frac{nq}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \sum_{j=1}^{q} \left(\ln\left(d_{ij}^*\right) - \ln\left(\sqrt{\sum_{k=1}^{s} (X_{ik} - Z_{jk})^2} \right) \right)^2$$

$$- \frac{1}{2\zeta^2} \left(\sum_{i=1}^{n} \sum_{k=1}^{s} X_{ik}^2 \right) - \frac{1}{2\kappa^2} \left(\sum_{j=1}^{q} \sum_{k=1}^{s} Z_{jk}^2 \right) - \ln(c)$$

$$= -\frac{nq}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \sum_{j=1}^{q} \left(\ln\left(d_{ij}^*\right) - \ln(d_{ij}) \right)^2 - \frac{1}{2\zeta^2} \left(\sum_{i=1}^{n} \sum_{k=1}^{s} X_{ik}^2 \right) - \frac{1}{2\kappa^2} \left(\sum_{j=1}^{q} \sum_{k=1}^{s} Z_{jk}^2 \right) - \ln(c). \tag{16}$$

We experimented with vague and informative priors for both the ζ and the κ and found that the results were essentially the same. In the Supplementary Appendix, we show the first and the second derivatives for equation (16).

Our unfolding example is the classic 1968 National Election Study feeling thermometers. A feeling thermometer asks the individuals to respond to a set of stimuli (political figures in this case) based on their subjective views of warmth toward them. The thermometer ranges from 0 to 100 degrees, with 100 indicating warm and very favorable feeling, 50 indicating neutrality toward the political figure, and 0 indicating that the respondent feels cold and very unfavorable toward the political figure. The 1968 feeling thermometers have been analyzed by Weisberg and Rusk (1970), Wang et al. (1975), Rabinowitz (1976), Cahoon, Hinich, and Ordeshook (1978), Poole and Rosenthal (1984), and Brady (1990), with the main focus on modeling the latent dimensions underlying the thermometers as well as testing theories of spatial voting.

In the NES 1968 survey, twelve political figures were included in the thermometer questions: George Wallace, Hubert Humphrey, Richard Nixon, Eugene McCarthy, Ronald Reagan, Nelson Rockefeller, President Johnson, George Romney, Robert Kennedy, Edmund Muskie, Spiro Agnew, and Curtis LeMay. There were 1673 respondents, and we included in our analysis the 1392 respondents who rated at least five of the twelve political figures.
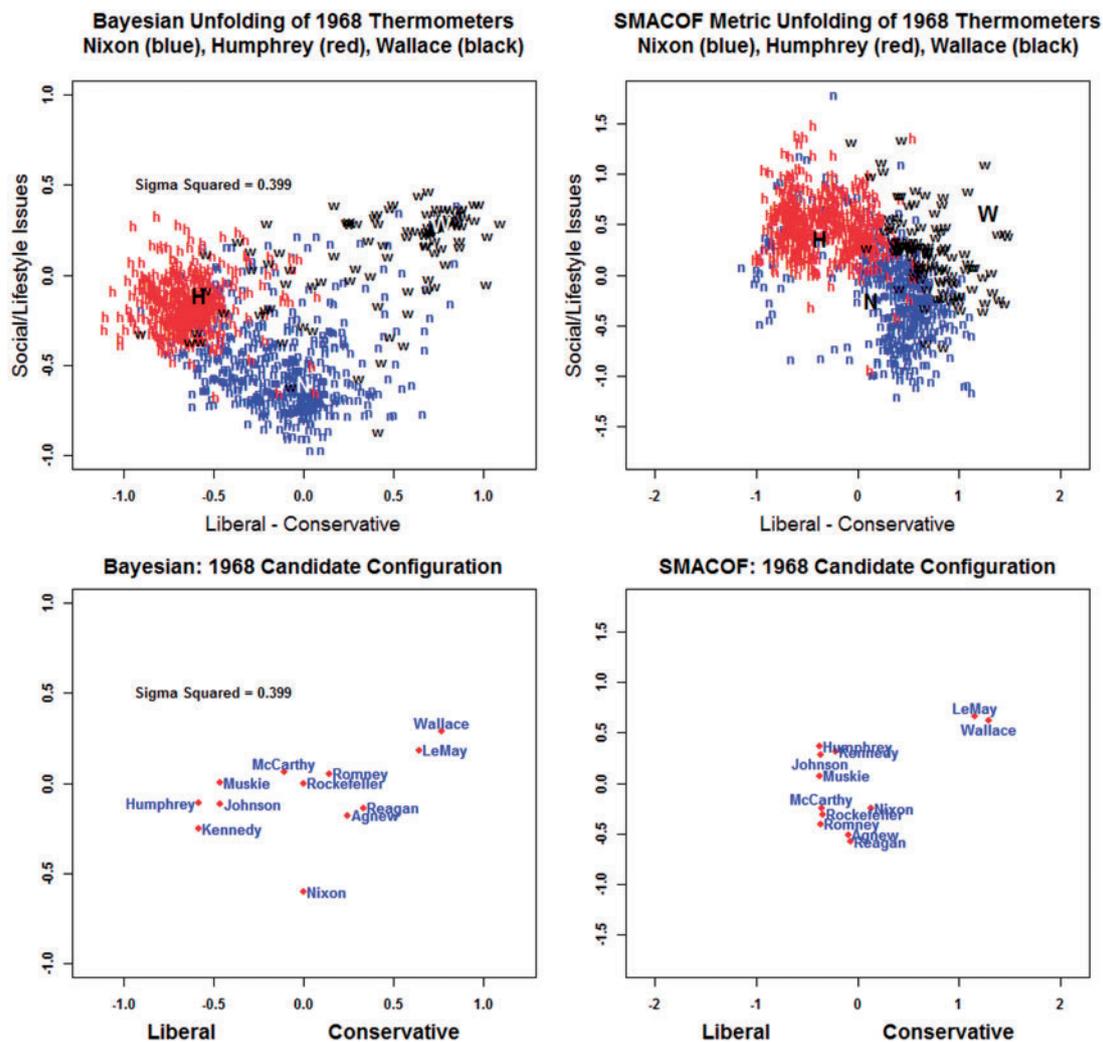

Fig. 4 The 1968 thermometer, Bayesian versus SMACOF unfolding.

We perform our analysis in two dimensions because previous analyses using optimization methods almost all find two dimensions in the data. We think this is due to the idiosyncratic noise in the thermometers (see Abrajano and Poole [2011] for a discussion) and valence effects (Merrill and Grofman 1999; Londregan 2000; Adams, Merrill, and Grofman 2005). A second dimension is picking up some of these effects and "smoothing" out the first dimension. Modeling valence effects are difficult, so we leave that for future work. In any event, our aim here is to show the advantages of our Bayesian approach. Namely, a properly designed Markov chain reveals much more information than simply the modes of a loss function.

Finding the optimal solution for the unfolding problem is challenging because of the number of parameters. We used limited-memory BFGS that can handle very large problems and is computationally efficient (see Liu and Nocedal [1989] for a discussion). In two dimensions, this required a search over a 2805 dimensional hyperplane. We set George Romney at the origin and Eugene McCarthy's second-dimension coordinate at zero to obtain identification. We checked the first derivatives (see the Supplementary Appendix) for the starting configuration to be sure that our points were located on a mode of the loss function (see equation (16)).

Using the respondent and candidate coordinates as targets, we were able to run a slice sampler on the 1968 data (Neal 2003). Since the ratio of respondents to the candidates is so large, we found that the method that worked the best was to first draw the respondent coordinates and then the
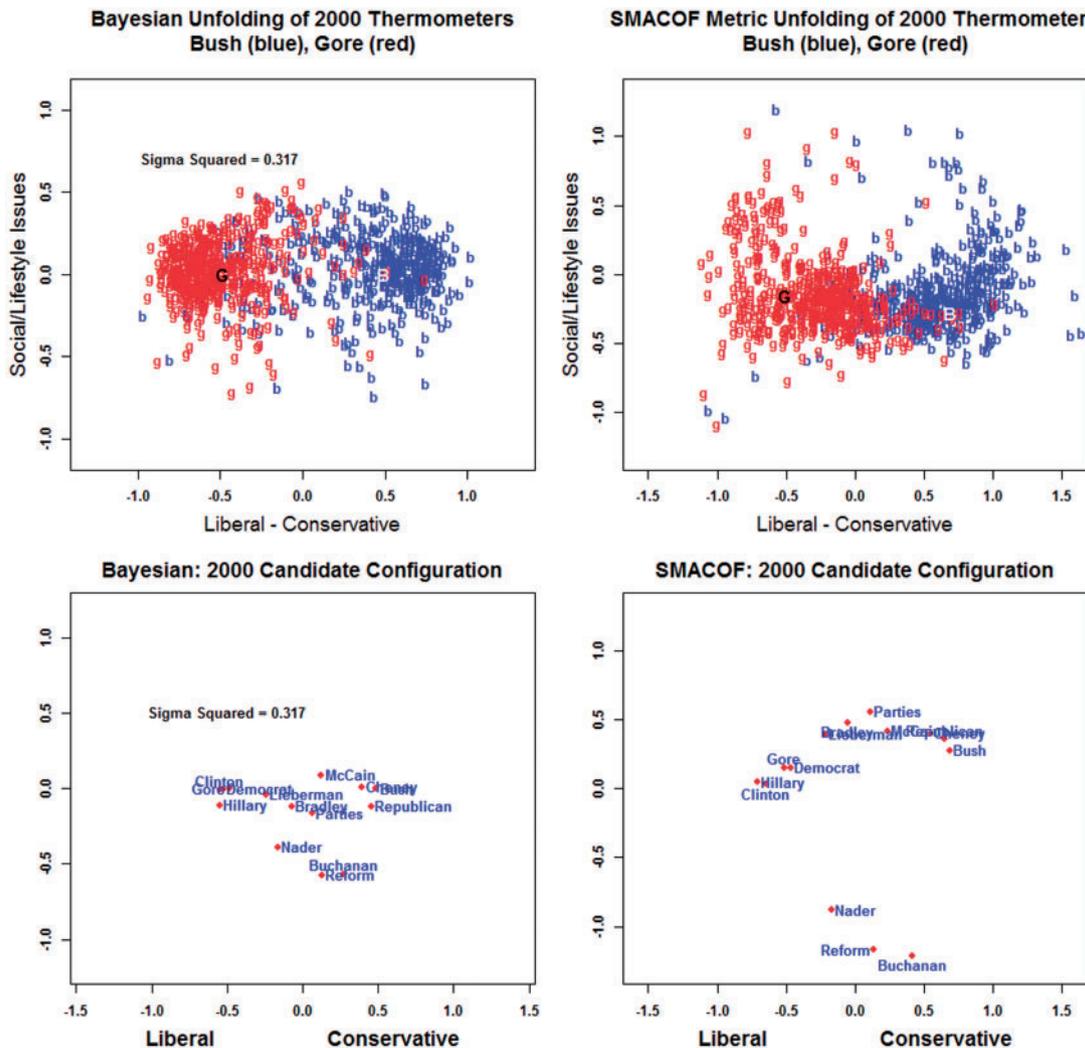

Fig. 5 The 2000 thermometer, Bayesian versus SMACOF unfolding.

candidate coordinates. We kept Romney at the origin, but we did not constrain any other points because we found that simply rotating the drawn configuration to the optimal configuration with Romney at the origin was simple and easy to implement (see the Supplementary Appendix for further detail). We ran our chain to 110,000 draws, with the first 10,000 as burn-in. Figure 4 shows the results.

The upper left panel of Fig. 4 shows the respondents along with tokens indicating the locations of Humphrey, Nixon, and Wallace. The lower left panel shows the candidate configuration. We display those respondents who indicated that they voted for Humphrey, Nixon, or Wallace as the tokens "h," "n," or "w," respectively. Humphrey, Nixon, and Wallace are located near where their voters are concentrated.

The upper right and the lower right panels of Fig. 4 show the results for an unfolding using SMACOF. Only Humphrey is located within where his voters are concentrated. In addition, the candidate configuration from SMACOF is bunched up save for Wallace and his running mate LeMay, who are pushed outward. These differences arise from the fact that the simple additive model (the true distance plus noise) as in equation (12) does not work well with the thermometer data. This is due to the fact that there are a lot of thermometer scores close to 100. The log-normal is a more realistic model of the data because the SE gets smaller as the distances get smaller.
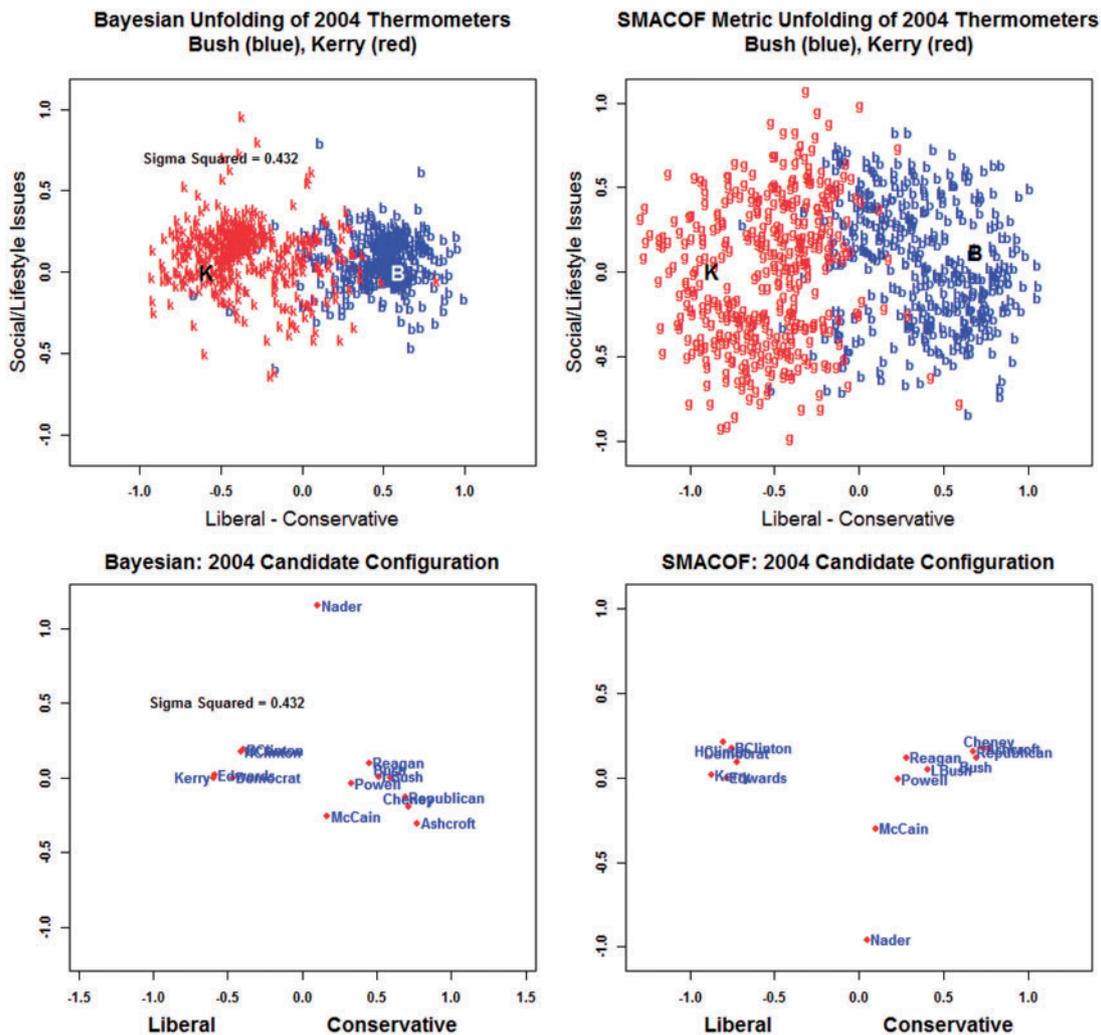

Fig. 6 The 2004 thermometer, Bayesian versus SMACOF unfolding.

1968 Candidate Contour Plots: Humphrey, Nixon, Wallace
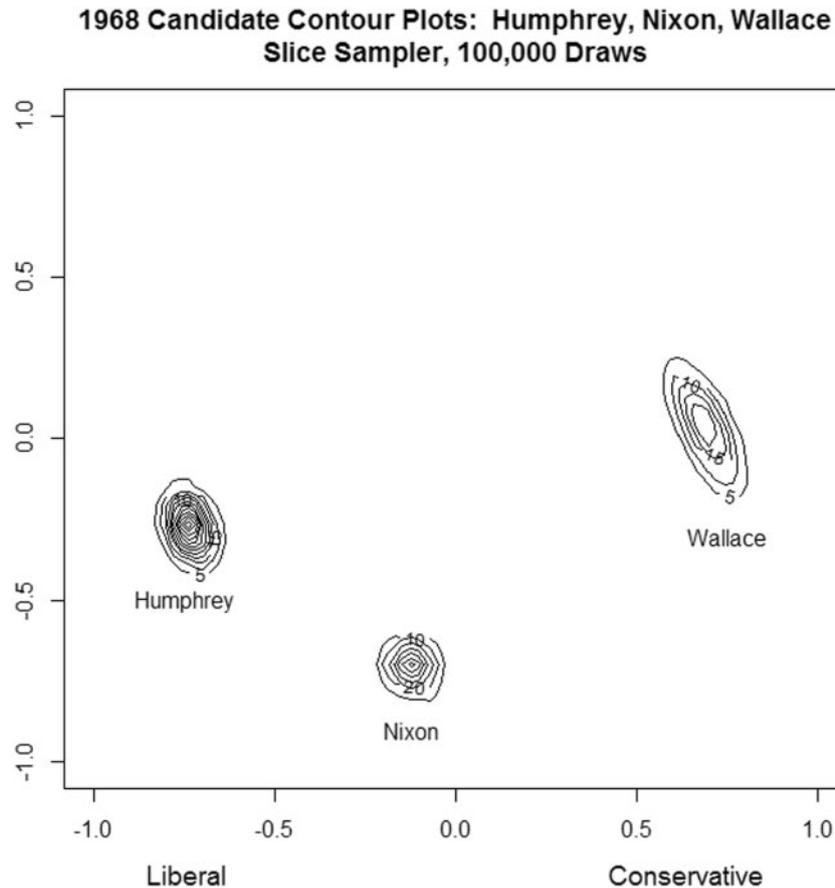## Slice Sampler, 100,000 Draws

Fig. 7 Major candidates, 1968 presidential election.

We also tested the MAP unfolding procedure developed by Hinich (2005). MAP uses a maximum likelihood factor analysis of the covariance matrix computed from the squared distances to recover the stimuli and then OLS using the stimuli to recover the respondent locations. Although MAP is computationally efficient, the configuration of respondents produced by MAP is not as clean as that produced by SMACOF. We consider SMACOF to be a better unfolding procedure. Hence, we show only comparisons with SMACOF.

Figures 5 and 6 show thermometer unfolding results for the 2000 and 2004 elections. Again, the Bayesian unfolding results place the major candidates among their supporters, and in these two elections the results are largely one-dimensional. The SMACOF results tend to spread the voters out more on the second dimension, but the location of the major presidential candidates is reasonable.

The candidates in the Bayesian unfoldings are precisely estimated. For example, in 1968, the largest SD was for George Wallace's second-dimension coordinate at 0.11. Figure 7 shows the bivariate densities for the three major 1968 presidential candidates. All are unimodal over the two-dimensional space.

The respondents were less precisely estimated. For example, Fig. 8 shows the distribution of the 100,000 draws for the coordinates of respondent number 2 in 1968. The second respondent was a young white male Democrat. He did not like Wallace, LeMay, Agnew, and Reagan (15, 30, 30, 30), but he was a little warmer toward Nixon and Romney (40, 40). He was less than enthused with President Johnson and Hubert Humphrey (50 and 60), but he really liked Robert Kennedy, Nelson Rockefeller, and Eugene McCarthy (97, 97, 85). His preferences roughly line up left to right but not entirely. This is reflected in the bimodal distribution of the draws. Most of the draws are in the lower mode around −0.7, −0.6, with a smaller number around −0.5, 0.1. This shows the advantage

1968 Respondent Contour Plot: Respondent 2
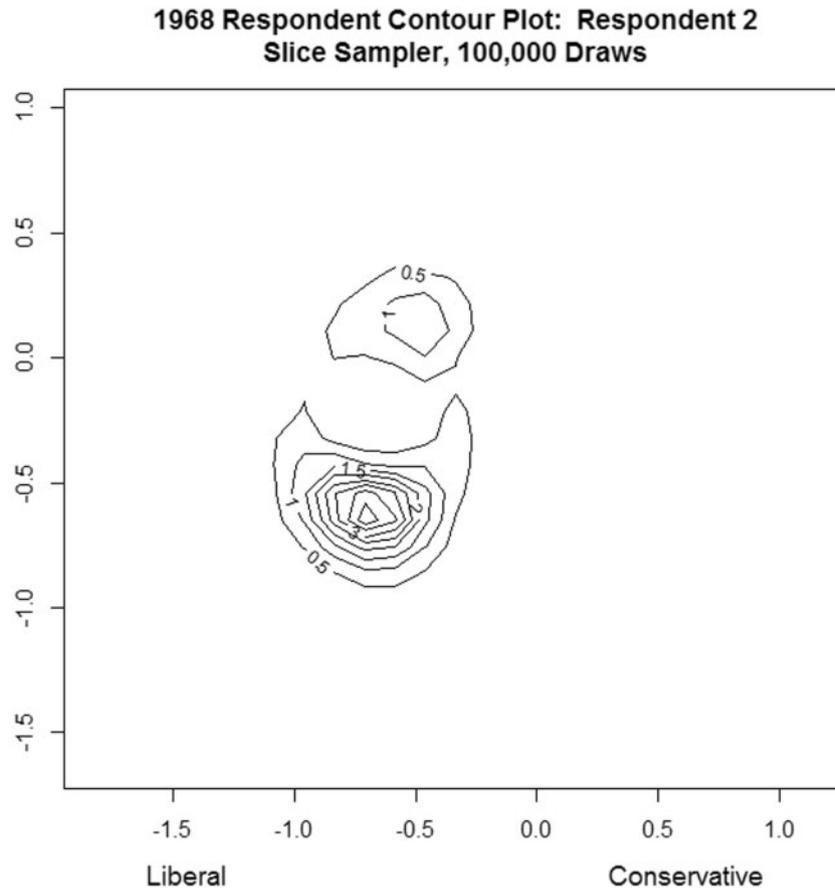## Slice Sampler, 100,000 Draws

Fig. 8 The 1968 NES, Respondent 2.

of a Bayesian approach. A mode finder (optimization method) will land on one of the two modes, whereas a Markov chain "illuminates" the entire distribution.

4   Conclusion

In this article, we have shown how to apply Bayesian methods to noisy ratio scale distances for both the classical similarities problem as well as the unfolding problem. Our approach combines the advantages of traditional mode finders and Bayesian MCMC. We use the mode finders to give us a target that identifies ("freezes") the posterior for the Markov chain generator.

Our unfolding example using the 1968 candidate thermometers shows the power of MCMC (made possible by the speed of modern computers) to illuminate complex distributions. Instead of modes with their associated SDs from the inverse Hessian, "painting," the entire posterior distribution allows us to show means and the complete distribution of the parameters.

Our results are preliminary. We deliberately kept our models simple because our aim was to revisit older problems using modern methods. We think the thermometers are an underutilized resource that potentially can reveal important information about individuals' utilities for political figures. Our aim here was simply to show a basic method that can be used as a springboard to more complex analyses.

References

Abrajano, Marisa, and Keith T. Poole. 2011. A method of linking surveys using affective signature with an application to ethnic groups in the U.S. In *Who Gets Represented*, eds. Peter Enns and Christopher Wlezien. New York: Russell Sage.

Adams, James, Samuel Merrill, and Bernard Grofman. 2005. *A unified theory of party competition: A cross-national analysis integrating spatial and behavioral factors*. New York: Cambridge University Press.

Bakker, Ryan, and Keith T. Poole. 2012. *Replication data for: Bayesian Metric Multidimensional Scaling*. http://hdl.handle.net/1902.1/18965 IQSS Dataverse Network [Distributor] V1 [Version]. Also available at http://voteview.com/Bakker_Poole_Bayesian_MDS.htm.

Bradlow, Eric T., and David C. Schmittlein. 2000. The little engines that could: Modeling the performance of World Wide Web search engines. *Marketing Science* 19(1):43–62.

Brady, Henry. 1990. Traits versus issues: Factor versus ideal-point analysis of candidate thermometer ratings. *Political Analysis* 2:97–129.

Cahoon, Lawrence S., Melvin J. Hinich, and Peter C. Ordeshook. 1978. A statistical multidimensional scaling method based on the spatial theory of voting. In *Graphical Representation of Multivariate Data*, ed. P.C. Wang. New York: Academic Press.

Clinton, Joshua D., Simon D. Jackman, and Douglas Rivers. 2004. The statistical analysis of roll call data: A unified approach. *American Political Science Review* 98:355–70.

De Leeuw, Jan. 1977. Applications of convex analysis to multidimensional scaling. In *Recent Developments in Statistics*, eds. J.R. Barra, F. Brodeau, G. Romier, and B. van Cutsem, 133–45. Amsterdam, The Netherlands: North-Holland.

———. 1988. Convergence of the majorization method for multidimensional scaling. *Journal of Classification* 5:163–80.

De Leeuw, Jan, and Heiser Willem. 1977. Convergence of correction-matrix algorithms for multidimensional scaling. In *Geometric Representations of Relational Data*, eds. J.C. Lingoes, E.E. Roskam, and I. Borg, 735–52. Ann Arbor, MI: Mathesis Press.

de Leeuw, Jan, and P. Mair. 2009. Multidimensional scaling using majorization: The R package smacof. *Journal of Statistical Software* 31(3):1–30, http://www.jstatsoft.org/v31/i03/.

DeSarbo, Wayne S., Joonwook Park, and Vithala R. Rao. 2011. Deriving joint space positioning maps from consumer preference ratings. *Marketing Letters* 22(1):1–14.

Eckart, Carl H., and Gale Young. 1936. The approximation of one matrix by another of lower rank. *Psychometrika* 1:211–8.

Enelow, James M., and Melvin J. Hinich. 1984. *The Spatial Theory of Voting*. New York: Cambridge University Press.

Fong, Duncan K. H., Wayne S. DeSarbo, Joonwook Park, and Crystal J. Scott. 2010. A Bayesian vector multidimensional scaling procedure for the analysis of ordered preference data. *Journal of the American Statistical Association* 105(490):482–92.

Gormley, Isobel C., and Thomas B. Murphy. 2006. A latent space model for rank data. *Statistical Network Analysis: Models, Issues, and New Directions: Lecture Notes in Computer Science*. New York: Springer.

Hinich, Melvin J. 2005. A new method for statistical multidimensional unfolding. *Communications in Statistics—Theory and Methods* 34:2299–310.

Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock. 2002. Latent space approaches to social network analysis. *Journal of the American Statistical Association* 97(460):1090–98.

Jacoby, William G. 1982. Unfolding the party identification scale: Improving the measurement an important concept. *Political Methodology* 8:33–60.

Kruskal, Joseph B. 1964a. Multidimensional scaling by optimizing a goodness of fit to a nonmetric hypothesis. *Psychometrika* 29:1–27.

———. 1964b. Nonmetric multidimensional scaling: A numerical method. *Psychometrika* 29:115–29.

Liu, Dong C., and Jorge Nocedal. 1989. On the limited memory BFGS method. *Mathematical Programming* 45(1–3):503–28.

Londregan, John B. 2000. Estimating legislators' preferred points. *Political Analysis* 8(1):35–56.

Martin, Andrew D., and Kevin M. Quinn. 2002. Dynamic ideal point estimation via Markov chain Monte Carlo for the U.S. Supreme Court, 1953–1999. *Political Analysis* 10:134–53.

Merrill, Samuel, III, and Bernard Grofman. 1999. *A unified theory of voting: Directional and proximity spatial models*. New York: Cambridge University Press.

Navarro, Daniel J., and Michael D. Lee. 2003. Combining dimensions and features in similarity-based representations. In *Advances in Neural Information Processing Systems*, eds. S. Becker, S. Thrun, and K. Obermayer, Vol. 15, 59–66. Cambridge, MA: MIT Press.

Neal, Radford M. 2003. Slice sampling. *Annals of Statistics* 31(3):705–67.

Nelder, John A., and Roger Mead. 1965. A simplex method for function minimization. *Computer Journal* 7:308–13.

Oh, Man-Suk, and Adrian E. Raftery. 2001. Bayesian multidimensional scaling and choice of dimension. *Journal of the American Statistical Association* 96(455):1031–44.

Okada, Kensuke, and Kazuo Shigemasu. 2010. Bayesian multidimensional scaling for the estimation of a Minkowski exponent. *Behavior Research Methods* 42(4):899–905.

Park, Joonwook, Wayne S. DeSarbo, and John Liechty. 2008. A hierarchical Bayesian multidimensional scaling methodology for accommodating both structural and preference heterogeneity. *Psychometrika* 73(3):451–72.

Poole, Keith T., and Howard Rosenthal. 1984. U.S. presidential elections 1968–1980: A spatial analysis. *American Journal of Political Science* 28:282–312.

———. 1997. *Congress: A political-economic history of roll call voting*. New York: Oxford University Press.

Pope, Jeremy C., and Shawn A. Treier. 2011. Reconsidering the great compromise at the federal convention of 1787: Deliberation and agenda effects on the senate and slavery. *American Journal of Political Science* 55:289–306.

Powell, Michael J. D. 1973. On search directions for minimization algorithms. *Mathematical Programming* 4:193–201.

Rabinowitz, George. 1976. A procedure for ordering object pairs consistent with the multidimensional unfolding model. *Psychometrika* 45:349–73.

Schönemann, Peter H. 1970. On metric multidimensional unfolding. *Psychometrika* 35:349–66.

Shepard, Roger N. 1962a. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika* 27:125–39.

———. 1962b. The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika* 27:219–46.

Torgerson, Warren S. 1952. Multidimensional scaling: I. Theory and method. *Psychometrika* 17:401–19.

———. 1958. *Theory and methods of scaling*. New York: Wiley.

Wang, Ming-Mei, Peter H. Schonemann, and Jerrold G. Rusk. 1975. A conjugate gradient algorithm for the multidimensional analysis of preference data. *Multivariate Behavioral Research* 10:45–80.

Weisberg, Herbert F., and Jerrold G. Rusk. 1970. Dimensions of candidate evaluation. *American Political Science Review* 64:1167–85.

Young, Gale, and Alston S. Householder. 1938. Discussion of a set of points in terms of their mutual distances. *Psychometrika* 3:19–22.