The Dimensions of Supreme Court Decision Making: Again Revisiting *The Judicial Mind*[*]

Andrew D. Martin
Department of Political Science
Washington University
Campus Box 1063
One Brookings Drive
St. Louis, MO 63130-4899
admartin@artsci.wustl.edu

Kevin M. Quinn
Department of Political Science and
Center for Statistics in the Social Sciences
Box 354322
University of Washington
Seattle, WA 98195-4322
quinn@stat.washington.edu

May 2, 2001

Abstract

At the heart of attitudinal and strategic explanations of judicial behavior is the assumption that justices have well-defined policy preferences. In the literature these preferences have been measured in a handful of ways, including using factor analysis and multidimensional scaling techniques (Schubert 1965, 1974), looking at past votes in a single policy area (Epstein et al. 1989), content-analyzing newspaper editorials at the time of appointment to the Court (Segal and Cover 1989), and recording the background characteristics of the justices (Tate and Handberg 1991). Scholars have used these measures successfully to explain behavior in some policy areas, including civil rights, but have been less successful in others, such as economics cases (Epstein and Mershon 1996). In this manuscript we employ Markov chain Monte Carlo (MCMC) methods to fit Bayesian measurement models of judicial preferences for the justices of the seventh Burger Court (1981-1985). In so doing, we simultaneously estimate a ideal point for each justice in a multi-dimensional space and the cutting hyperplane for each non-unanimous case. In general, estimating these models for the Supreme Court is quite difficult due to the small number of justices. The models we propose are thus developed to address aspects of this 'micro-committee problem.' The Bayesian approach allows us to include additional information about the cases being studied in a straightforward manner, and also allows us to gauge the uncertainty of each estimate by summarizing the posterior density. We present results for several models, compare the estimated bliss points across models, and utilize the findings to study two cases of substantive importance: *Garcia v. San Antonio Metropolitan Transit Authority* (469 US 528) and *Dixson v. United States* (465 US 482).

[*]Paper prepared for presentation at the Midwest Political Science Association Annual Meeting, Chicago, Illinois, April 19-22, 2001, and at the Modelling U.S. Politics Conference hosted by the Center in Political Economy at Washington University, May 11-13, 2001. We would like to thank Lee Epstein, Steve Van Winkle, and Scott Comparato for useful discussions along the way. All analyses were performed on Linux or BSD UNIX workstations using C++ and R code written by the authors. We gratefully acknowledge financial support from the Center in Political Economy at Washington University. Data for replication can be obtained from the Supreme Court Database website at Michigan State University: http://www.ssc.msu.edu/~pls/pljp/sctdata.html. Documented C++ code for these models will be available in late Summer 2001 at the authors' websites.

Introduction

Over forty years ago Glendon Schubert began applying psychometric models to the study of the Supreme Court (Schubert 1959). These, models, developed by Louis Thurstone, Charles Guttman, Clyde Coombs, and others, are based on multiple factor analysis, and had been previously applied to the study of personality, opinion, and ideology. Schubert was substantively building on the work of Pritchett (1948), who recognized that one could perform rigorous empirical analyses derived from the legal realism approach. In *The Judicial Mind*, Schubert (1965) presents a comprehensive look at the attitudes and ideologies of all Supreme Court justices serving from the end of the Roosevelt era to the beginning of the Johnson administration. His i-points (what we would now term ideal points) and j-points (currently called a cutpoint or cleavage line), along with his C scale (political liberalism) and E scale (economics), became an important part of the judicial politics vernacular, and let to a "behavioral revolution" in the study of Supreme Court decision making.

A decade after *The Judicial Mind* was published Schubert returned to the subject in *The Judicial Mind Revisited* (1974). He explains the reason for this: "One major weakness of the earlier study was the relative crudity of the techniques for data analysis then available, even at the avant-garde establishment where my work was done" (p. 2). Improvements in available methodology is what drove Schubert to revisit the subject, and these improvements are what motivates us to again revisit the judicial mind. Over the last twenty-five years there has been a growth in the psychometric and educational testing literature that allow us to get a better picture of decision making on the Supreme Court. Additionally, studying an institution with only nine members poses important methodological challenges not fully addressed in the statistical literature.

The purpose of this paper is to present a handful of models that can be used to estimate ideal points of Supreme Court justices. This is well-tread methodological ground; in every decade since the 1950s, scores of articles have been published that estimate ideal points for legislatures and courts using various techniques. Perhaps the best example is the work by Keith Poole and Howard Rosenthal, who study the entirety of roll call voting in the United States House and Senate (Poole and Rosenthal 1991a, 1997). Not only have their measures informed a generation of Congress scholars, but the computational tools brought to the problem were quite innovative. As detailed below, we depart from the existing literature in two fundamental ways. First, we develop models better suited to studying small deliberative bodies, such as the Supreme Court. We agree with Londregan (2000a) that the 'micro-committee problem' limits the value of extant approaches. Most of the variants of the standard item response model we present are specifically geared to solving particular aspects of this methodological challenge. Second, we adopt a Bayesian approach. But for the development of Markov chain Monte Carlo (MCMC) methods in the late 1980s and early 1990s, the models we propose would have been intractable. Others have performed Bayesian inference for standard item response models (Albert 1992; Patz and Junker 1999) and item response models applied to voting data (Clinton et al. 2000), but to our knowledge no one has developed models and classes of prior distributions designed to address the problems encountered in situations with a small number of subjects. We find this interesting since, as Londregan (2000a) points out, including additional prior knowledge is the only way around the micro-committee problem.

This paper proceeds as follows. In the following section we review other approaches scholars have employed to measure judicial preferences. This is followed by a discussion of research design, and an introduction to our modeling approach. Next, we discuss issues of identification, and propose partial solutions to the micro-committee problem. Thereafter we present results from one-dimensional and two-dimensional models for the seventh Burger Court, and compare the findings of each model. We also use the model to compute substantively interesting quantities of interest, including the posterior probability distribution of the median justice, and to study two cases: *Garcia v. San Antonio Metropolitan Transit Authority* (469 US 528) and *Dixson v. United States* (465 US 482). The final section concludes with a discussion of more comprehensive solutions to the micro-committee problem and the broader research project.

Measuring Judicial Preferences

Over the last fifty years many scholars have devoted a substantial amount of time estimating ideal points for Supreme Court justices. The reason for this work is clear. Whether one adopts an attitudinal (Segal and Spaeth 1993) or strategic (Eskridge 1991; Epstein and Knight 1998) approach to explain the behavior of justices or the policy outputs of the institution, one requires a measure of the ideal policy position of each justice (or the Court, which is typically measured using the median justice). Here we remain agnostic as to whether justices behave aptitudinally or strategically.[1] Like many previous scholars, we are interested in estimating the policy preferences of justices. Unlike many others, we are not interested in creating a set of preference measures that can be simply plugged into regression models of various sorts as either left-hand-side or right-hand-side variables. Indeed, we believe that such approaches are often inappropriate since the assumptions often used to construct the measures of preferences are inconsistent with aspects of the regression models used by future researchers. Instead, we believe that if one wants to make causal inferences about the effects of policy preferences, case characteristics, and other relevant independent variables on Supreme Court voting behavior, one must specify a structural model derived from theory and fit the resulting model. This is what we pursue in this paper. An inherent part of this exercise is operationalizing the issue space over which justices are making decisions. One important questions is: Over how many dimensions are justices voting?

The first genre of research relating to ideal point estimation grew from the work of Schubert (1959). In these approaches, one uses the matrix of votes to estimate an ideal point using various data summarization techniques, including multidimensional scaling, Guttman (cumulative) scaling, and factor analysis. This research, firmly based in behavioral political science, produced a vast amount of knowledge regarding preferences of individual justices, and the doctrinal interrelationships between cases. Schubert (1965) finds two primary dimensions – political and economic

[1] The attitudinal model is based on psychological stimulus-response models, while strategic approaches are based on economic rational actor models. Both can be viewed in terms of the spatial voting model (Downs 1957; Enelow and Hinich 1984), where each actor has an ideal point and votes for the policy alternative closest to that point. For expositional economy, we employ the parlance of the spatial voting model throughout this paper.

liberalism – that explain votes on the merits for the Court from the late 1940s to the early 1960s. He also finds evidence of three "minor scales" which are salient in only a subset of terms. Nine years later, Schubert (1974) again finds that these two dimensions structure behavior through the Vinson and Warren Courts. Rohde and Spaeth (1976) take a similar approach, and find three primary values that structure behavior for the Warren Court and first five terms of the Burger Court: freedom, equality, and what they term "New Dealism." These three values explain over 85% of votes on the merits during this time period. Spaeth (1990) extends this analysis through the entirety of the Burger Court, and looks specifically at many sub-scales relating to civil liberties. He demonstrates argues that civil liberties cases are not all fungible, and that these sub-scales are correlated highly with the equality dimension found by Rohde and Spaeth (1976).

A similar approach relies on using past behavior in a policy area to explain votes in the same policy area in the future. Epstein et al. (1989) look at final case decisions in all criminal justice cases before the Supreme Court between 1946 and 1986, and model the percentage of criminal cases that the Court supported the criminal rights position. To gauge preferences of the Court they simply use the percentage of pro-criminal rights decisions in the previous term. Similarly, Segal and Spaeth (1989) report the fraction of liberal decision on civil liberties and economic cases for all justices serving between 1953 and 1985. They find a great deal of variance within and between these two most salient dimensions.

There are a handful of drawbacks to these approaches. First, they are not based on parametric statistical models and are best seen as data summarization techniques. The uncertainty in the measures cannot be gauged, and one can neither test for dimensionality nor compare various models in a principled fashion. This is not to say that these approaches are inherently problematic, but as we argue below, one can do better. Second, scholars have argued that because these measure are simple summaries of past votes, they cannot be used to explain individual votes.

To alleviate this circularity problem, scholars have developed a handful of measures not based on the votes of the justices. One literature uses social background characteristics to measure ideal points. Tate and Handberg (1991) look at the long-term behavior of Supreme Court justices in civil rights and economics cases from 1916 to 1988. Their personal attribute model asserts that political factors, salient social cleavages, personal origins, and career experience can be used to explain voting behavior on the Court. The results suggest that for both issue areas, political factors (the partisanship of the justice, and the intentions of the appointing president), and whether justices had previous prosecutorial experience are the key explanatory factors of judicial behavior. Theoretically, however, it is difficult to believe that these demographic characteristics are explanatory, and thus can be applied out-of-sample.

Perhaps the most novel preference measure in the judicial politics literature was offered by Segal and Cover (1989). In response to the criticism that the evidence for the attitudinal model was merely using votes to explain votes, Segal and Cover developed an exogenous measure of judicial preferences by systematically analyzing newspaper editorials before confirmation hearings. These measures were computed for all justices serving from Earl Warren to Anthony Kennedy. While these have been applied in many different ways, the authors argue that they should only be used to study aggregate votes on the civil liberties cases, thus limiting their applicability. Segal et al. (1995) updates these scores by adding Bush appointees, and backdates the scores by computing them for seven Roosevelt and the four Truman nominees. Epstein and Mershon (1996) perform a "methodological audit" of this measure, and demonstrate that they should *only* be used for their designed purpose: explaining aggregated civil liberties votes. It is interesting to note that these scores explain less than 20% of the variance in economics cases, the other salient dimension documented in previous scholarship. Not only does this unidimensionality problem plague Segal/Cover scores, but there is also an assumption that justices' preferences are constant over time. Epstein et al. (1998) take a dynamic approach, and find that Baum-corrected (Baum 1988) percent liberal civil liberties decisions varies for certain justices over time. Thus, estimating a single ideal point for each justice, such as the Segal/Cover approach, is inappropriate.

While these approaches are highly creative, we feel that a better approach to explaining votes as a function of policy preferences (and other potential causal factors) is to posit a theoretical model of judicial decision-making and then directly operationalize it into a statistical model. Such a structural model has the property that if the theoretical model is correct, thus implying that the posited statistical assumptions hold, then such a model can explain votes in terms of preferences even though preferences are treated as latent variables[2].

We glean a number of additional lessons from the literature. First, there are likely two, and perhaps more, dimensions that structure behavior on the Court. At a minimum, we expect to find a political liberalism / civil rights dimension that structures behavior as well as an economics dimension. Second, static measures are inappropriate; that is, individual preferences can change over time. We do not address this here, although as discussed in the conclusion, our modeling strategy can be parameterized to address preference change. We also choose to follow the lead of Schubert and use votes to measure ideal points.

Modeling Supreme Court Decision Making

Our focus in this paper is the seventh natural Burger Court, which served from 1981 to 1985. We treat these data as if they come from a single cross-section. Thus, we assume that during this time period, each justice has a fixed ideal point.[3] We our obtain data from *The United States Supreme Court Database* (Spaeth 1999). We select all non-unanimous cases that are: formally decided with written opinion, decided *per curiam*, decrees, decided by an equally dived votes, unsigned *per curiam* cases, and judgements of the Court.[4] This results in votes for $J = 9$ justices: Blackmun, Brennan, Burger, Marshall, O'Connor, Powell, Rehnquist, Stevens, and White. During these terms, the Court decided $K = 483$ cases. As discussed below, we use data augmentation to

[2]This focus on a unified structural model of voting and preference estimation is also found in Clinton et al. (2000).

[3]One key tenet of the strategic approach is that justices may behave in a non-sincere fashion. If it is the case justices are behaving strategically, their revealed preferences may not be the same as their sincere preferences. Because we look at a single natural court serving during the first Reagan administration, the effect of intra-Court or inter-institutional (i.e., separation of powers) considerations would likely be constant. Thus, if one believes the attitudinal model, our estimates could be viewed as true ideal points. If one believes the strategic model, our estimates should be viewed as revealed preferences.

[4]The unit of analysis is the case citation (ANALU=0). We select cases where DEC_TYPE equals 1, 2, 4, 5, 6, or 7, and drop all cases when VOTE equals 40, 50, 60, 70, 80, and 90. It would be possible to include unanimous decisions in our analysis if one is willing to fix *a priori* the discrimination and difficulty parameters for these cases.

deal with missing data problems. Powell missed the most votes (twenty-nine), and Burger missed the fewest (two). The data matrix is thus a 483×9 matrix of zeros, ones, and missing values.

We are interested in modeling the decisions of justices $j = 1, \ldots, J$ deciding cases $k = 1, \ldots, K$ in a D-dimensional issue space. Our assumption is that each justice's vote is an expressive action and depends only on the value the justice attaches to the policy positions of the status quo and the policy alternative. Put another way, a justice will vote to affirm the decision of the lower court if the utility the justice attaches to the status quo is greater than the utility the justice attaches to the alternative, regardless of the expected actions of the other actors.

To operationalize this model we begin by writing down random utility functions. Let

$$u_{jk}^{(a)} = -\|\boldsymbol{\theta}_j - \mathbf{x}_k^{(a)}\|^2 + \delta_{jk}^{(a)}$$

be the utility to justice i of voting to affirm on case j, and

$$u_{jk}^{(r)} = -\|\boldsymbol{\theta}_j - \mathbf{x}_k^{(r)}\|^2 + \delta_{jk}^{(r)}$$

be the utility to justice j of voting to reverse on case k, where $\boldsymbol{\theta}_j$ is justice j's ideal point in the D-dimensional issue space, $\mathbf{x}_k^{(a)}$ is the location of the policy under an affirmative vote, $\mathbf{x}_k^{(r)}$ is the location of the policy under a reversal, and $\delta_{jk}^{(a)}$ and $\delta_{jk}^{(r)}$ are Gaussian disturbances with zero mean and variances τ_a^2 and τ_r^2 respectively.

Given this spatial model, justice j will vote to affirm on case k when $u_{jk}^{(a)} > u_{jk}^{(r)}$ or equivalently when $u_{jk}^{(a)} - u_{jk}^{(r)} > 0$. Let z_{jk} be the difference between $u_{jk}^{(a)}$ and $u_{jk}^{(r)}$. We can write and simplify this utility differential z_{jk} as follows [5]:

$$\begin{aligned}
z_{jk} = u_{jk}^{(a)} - u_{jk}^{(r)} &= -\|\boldsymbol{\theta}_j - \mathbf{x}_k^{(a)}\|^2 + \delta_{jk}^{(a)} + \|\boldsymbol{\theta}_j - \mathbf{x}_k^{(r)}\|^2 - \delta_{jk}^{(r)} \\
&= -\left[\boldsymbol{\theta}_j - \mathbf{x}_k^{(a)}\right]' \left[\boldsymbol{\theta}_j - \mathbf{x}_k^{(a)}\right] + \delta_{jk}^{(a)} + \left[\boldsymbol{\theta}_j - \mathbf{x}_k^{(r)}\right]' \left[\boldsymbol{\theta}_j - \mathbf{x}_k^{(r)}\right] - \delta_{jk}^{(r)} \\
&= -\boldsymbol{\theta}_j'\boldsymbol{\theta}_j + 2\boldsymbol{\theta}_j'\mathbf{x}_k^{(a)} - \mathbf{x}_k^{'(a)}\mathbf{x}_k^{(a)} + \delta_{jk}^{(a)} + \boldsymbol{\theta}_j'\boldsymbol{\theta}_j - 2\boldsymbol{\theta}_j'\mathbf{x}_k^{(r)} + \mathbf{x}_k^{'(r)}\mathbf{x}_k^{(r)} - \delta_{jk}^{(r)} \qquad (1) \\
&= \left[\mathbf{x}_k^{'(r)}\mathbf{x}_k^{(r)} - \mathbf{x}_k^{'(a)}\mathbf{x}_k^{(a)}\right] + 2\boldsymbol{\theta}_j' \left[\mathbf{x}_k^{(a)} - \mathbf{x}_k^{(r)}\right] + \left[\delta_{jk}^{(a)} - \delta_{jk}^{(r)}\right] \\
&= \alpha_k + \boldsymbol{\beta}_k'\boldsymbol{\theta}_j + \varepsilon_{jk}
\end{aligned}$$

Here α_k is a scalar, $\boldsymbol{\beta}_k$ is a $D \times 1$ column vector, and $\boldsymbol{\theta}_j$ is a $D \times 1$ column vector. Given this model, justice j will vote to affirm on case k when $z_{jk} > 0$.

The Basic Item Response Model

The process of translating the spatial voting model above (presented in Equation 1) into a statistical model begins by noting that our observed data $\mathbf{Y}$ takes the form of a $J \times K$ matrix containing the votes of the J justices on the K cases under consideration. We code a vote to affirm as a one, a vote to reverse as a zero, and no reported vote as a missing value. For observed votes, we assume that:

$$y_{jk} = \begin{cases} 1 & \text{if} \quad z_{jk} > 0 \\ 0 & \text{if} \quad z_{jk} \leq 0 \end{cases} \qquad (2)$$

where,

$$z_{jk} = \alpha_k + \boldsymbol{\beta}_k'\boldsymbol{\theta}_j + \varepsilon_{jk} \qquad \varepsilon_{jk} \sim \mathcal{N}(0, 1). \qquad (3)$$

Note that we have fixed the variance of ϵ_{jk} to one since this variance and the other model parameters are not separately identified in the likelihood. This is a typical assumption; e.g., one makes this assumption for the standard probit model for a dichotomous dependent variable.

To form a probability model, we assume that:

$$y_{jk}|\alpha_k, \boldsymbol{\beta}_k, \boldsymbol{\theta}_j \overset{iid}{\sim} \mathcal{B}ernoulli(\pi_{jk}).$$

[5] Clinton et al. (2000) derive the same expression in the context of a model of legislative voting and go on to show the link with standard item response models.

Taken together, Equations 2 and 3 imply that $\pi_{jk} = \Phi(\alpha_k + \boldsymbol{\beta}_k'\boldsymbol{\theta}_j)$ where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Thus, the sampling density of the observed $\mathbf{Y}$ is:

$$f(\mathbf{Y}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{k=1}^{K}\prod_{j=1}^{J} \Phi(\alpha_k + \boldsymbol{\beta}_k'\boldsymbol{\theta}_j)^{y_{jk}} \left[1 - \Phi(\alpha_k + \boldsymbol{\beta}_k'\boldsymbol{\theta}_j)\right]^{1-y_{jk}}$$

Bayesian inference for the justices ideal points (the $\boldsymbol{\theta}_j$s) and the case parameters (the α_ks and the $\boldsymbol{\beta}_k$s)[6] proceeds by summarizing the posterior density given by:

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{Y}) \propto f(\mathbf{Y}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})p(\boldsymbol{\alpha})p(\boldsymbol{\beta})p(\boldsymbol{\theta})$$

where $p(\boldsymbol{\alpha})$, $p(\boldsymbol{\beta})$, and $p(\boldsymbol{\theta})$ are densities representing our prior beliefs about these parameters. Our various choices of prior distributions are discussed in the following section. The Markov chain Monte Carlo (MCMC) algorithms used to sample from and summarize the posterior distributions of these models are detailed in Appendices A-B.

Mixtures of Item Response Models

While the models discussed above allow the underlying issue space to be multidimensional, these models implicitly assume that each case has policy consequences on all D issue areas. Furthermore, this class of models assume each issue dimension is equally salient for each case. The implications of this are substantively problematic. One would have to believe, for example, that the economic aspects of *City of Akron v. Akron Center for Reproductive Health* (462 U.S. 416) would be equally important to each justice as the civil liberties consequences of the decision. Clearly this goes against much of the conventional wisdom regarding Supreme Court decision making. Another potential problem with the multidimensional variants of the models discussed above is that such models suggest a degree of policy instability that is not seen on the Court. The chaos theorems (e.g. McKelvey 1979) would predict that if the issue space is two-dimensional or greater, we would expect to see majority cycling. But cycling will only occur if single cases address multiple dimensions. If it is the case that each case only impacts a single dimension, cycling will not occur (Shepsle 1979).

[6]In the education testing literature, where these models were developed, these parameters are called the difficulty and discrimination parameters respectively.

In this section, we derive a statistical model that allows the salience of each issue dimension to vary across cases. Theoretically this approach implies a degree of policy stability more in line with what is seen on the Court. One could argue that precedent constrains the issues under consideration on each case, and thus we would expect each case to have policy implications on only one dimension.

Once again, we assume a D-dimensional issue space. However, we now assume that there is a probability ρ_{kd} that case k is decided solely on the basis of issue dimension $d \in \{1, \ldots, D\}$. More formally, we assume that the sampling density of $\mathbf{Y}$ is given by:

$$f(\mathbf{Y}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\rho}) = \prod_{k=1}^{K} \sum_{d=1}^{D} \left\{ \rho_{kd} \prod_{j=1}^{J} \Phi(\alpha_k + \boldsymbol{\beta}_k' \boldsymbol{\theta}_{jd})^{y_{jk}} \left[1 - \Phi(\alpha_k + \boldsymbol{\beta}_k' \boldsymbol{\theta}_{jd})\right]^{(1-y_{jk})} \right\}$$

where $\sum_{d=1}^{D} \rho_{kd} = 1$ for all k. Note that we allow the justices ideal points to vary across the dimensions of the issue space. The posterior density of this model is:

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\rho}|\mathbf{Y}) \propto f(\mathbf{Y}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\rho}) p(\boldsymbol{\alpha}) p(\boldsymbol{\beta}) p(\boldsymbol{\theta}) p(\boldsymbol{\rho})$$

Once again, our choices of prior distributions are discussed in the following section, and Markov chain Monte Carlo algorithms used to sample from the posterior distribution are detailed in Appendix C.

The simplest version of this model assumes that the probability of a case falling on a given issue dimension is constant across all cases. In other words, that $\rho_{kd} = \rho_{k'd} = \rho_d$ for all k and k'. Another choice of parameterization (for the $D = 2$ case) is to assume that:

$$\rho_{k2} = \frac{\exp(\mathbf{w}_k' \gamma)}{1 + \exp(\mathbf{w}_k' \gamma)} \qquad \text{and} \qquad \rho_{k1} = 1 - \rho_{k2}$$

where $\mathbf{w}_k$ is a $(Q \times 1)$ vector of covariates specific to case k that are believed to be predictive of the issue dimension that case was decided upon. In other words, one uses a logit model to predict which issue dimension a particular case falls. The ρ_{k1} and ρ_{k2} parameters tell us the posterior probability that case k falls on either dimension one or dimension two respectively.[7] We discuss estimation for the mixture model with covariates in Appendix D.

[7] A multinomial logit parameterization can be used when $D > 2$.

Identification and the 'Micro-Committee Problem'

It is well known that item response models suffer from identification problems (see, for example, Albert 1992; Johnson and Albert 1999). The first problem is called the 'scale invariance problem.' The parameters of interest $\boldsymbol{\alpha}, \boldsymbol{\beta},$ and $\boldsymbol{\theta}$ are identified only up to an arbitrary scale factor. Thus, one must anchor the scale on which ideal points are measured, just like one would fix a scale to measure any quantity.[8] In the Bayesian context, on fixes the scale invariance problem using a prior probability on the bliss points $\boldsymbol{\theta}$. The standard prior distribution for the ideal points $\boldsymbol{\theta}_j$s is to assume $\boldsymbol{\theta}_j \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I})$ for all justices $j = 1, \ldots, J$; i.e., one assumes independent standard Normal priors on the ideal points. We discuss an alternative prior below.

An additional identification problem called 'rotational' invariance. For the one-dimensional case ($D = 1$), multiplying all of the model parameters by negative one would not change the value of the likelihood function. Substantively, the model cannot determine what direction is liberal or conservative; i.e., should Thurgood Marshall be given a large positive or large negative score? It makes no substantive difference whether we estimate a liberalism or a conservatism scale. We also remedy this problem by using prior distributions. For the one-dimensional case, we assign zero prior probability that Marshall falls above zero. This forces us to estimate a conservatism measure, and thus fixes rotational invariance. For the two-dimensional case ($D = 2$), there are four posterior modes of equal probability. Thus, one has to impose three constraints to ensure that we are exploring a single mode. In this case, the space can be rotated and yield the same substantive results.[9] For each model we estimate below, we discuss what additional constraints we impose *a priori* to ameliorate this problem.

So far we have discussed prior distributions for the bliss points $\boldsymbol{\theta}$, but it is also necessary assign prior probabilities to the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ parameters. The standard approach is to assume that these are

[8]For example, it matters not whether one measures distance in miles or millimeters; the only difference between the measures is a scale factor.

[9]Londregan (2000a) discusses an additional identification problem relating to likelihood based inference. Due to granularity of the voting data, estimates of the ideal points are not likelihood-identified. He proposes a solution using agenda procedures. This problem does not plague the Bayesian approach. Additionally, modeling agendas is not a straight forward exercise for the Supreme Court.

drawn from Normal distributions. In other words:

$$\begin{bmatrix} \alpha_k \\ \boldsymbol{\beta}_k \end{bmatrix} \sim \mathcal{N}_{D+1}(\mathbf{b0}, \mathbf{B0})$$

In most applications, estimating these parameters is straight forward. There are many actors under study (whether students in the educational testing literature, or legislators in political science). However, when studying the Supreme Court, it becomes clear that estimating these parameters will be quite difficult because $J = 9$. Indeed, for each case, we must estimate $(D + 1)$ case-specific parameters using essentially a probit model with $J = 9$ observations. Our experience demonstrates that reliably estimating these parameters, using either frequentist or Bayesian methods, is impossible. This 'micro-committee problem' is not unique to studying the Supreme Court. Londregan (2000b) studies legislative committees in Chile, and uses information about the agenda process to solve the problem. We briefly discuss how a variant of this approach might be applied to the Court in the conclusion. In the remainder of this section, we detail two solutions to this methodological problem.

Hierarchical Priors and Covariates

Instead of assuming that the case-specific parameters are drawn from a common multivariate Normal distribution with mean $\mathbf{b0}$ and variance $\mathbf{B0}$, we can employ case-specific covariates to model the variance in these parameters across cases (for a related approach, see Bradlow 2000). It is likely that cases in similar issue areas, for example, will share similar α_k and $\boldsymbol{\beta}_k$ parameters. Instead of shrinking all estimates to zero, it would be better to shrink cases in different issue areas to different means. In so doing, we would allow cases to borrow strength from similar cases, and thus get more precise estimates of the case-specific parameters.

To accomplish this we employ a standard hierarchical linear model. For the α_k parameters, we assume:

$$\alpha_k \sim \mathcal{N}(\mathbf{w}_k' \boldsymbol{\xi}_\alpha, \sigma_\alpha^2)$$

Where $\mathbf{w}_k$ is a $(Q \times 1)$ vector of case-specific explanatory variables, $\boldsymbol{\xi}_\alpha$ is a $(Q \times 1)$ vector of parameters, and σ^2_α is a scalar variance parameter. Similarly, for each of the D elements of the $\boldsymbol{\beta}_k$ vector, we assume:

$$\beta_{k,d} \sim \mathcal{N}(\mathbf{w}_k' \boldsymbol{\xi}_{\beta,d}, \sigma^2_{\beta,d}) \quad \text{for all } d = 1, \ldots, D$$

In practice, any case-specific covariates can be used. Here we use a vector of dummy variables which signify which issue area, coded by Spaeth (1999), a specific case falls on.[10] Our expectation is that using this hierarchical prior will allow us to get much better estimates of the case-specific parameters, which are a necessary component to understanding decision making on each case.

A Prior Distribution on the Ideological Ordering of the Justices

Since, as noted above, the scale of the justices' ideal points is arbitrary, it is often difficult for researchers to specify reasonable prior distributions for these ideal points. As noted above, the default prior distribution for the $\boldsymbol{\theta}_j$s in these models is to assume $\boldsymbol{\theta}_j \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}), j = 1, \ldots, J$. The problem with this prior is that when information exists as to the order of the $\boldsymbol{\theta}_j$s, the most extreme ideal points will be overly shrunk towards zero. Another problem is that the symmetry of this prior results in a posterior that is multimodal.

An alternative is to specify a prior distribution over the orderings of the ideal points which, with some minor additional assumptions, is sufficient to implicitly define a prior density over the ideal points themselves. For simplicity, we begin with the case of unidimensional θ_j, and assume there exist bounds on the possible values of each θ_j. Specifically, we assume $\theta_j \in [-\lambda, \lambda], \quad j = 1, \ldots, J$. This implies that $\boldsymbol{\theta} \in \boldsymbol{\Theta} \equiv [-\lambda, \lambda]^N$. Since there are J ideal points, there are $J!$ possible orderings of the ideal points. Each of these $J!$ orderings corresponds to a subset of $\boldsymbol{\Theta}$. We denote the subset of $\boldsymbol{\Theta}$ that corresponds to ordering l as S_l. Note that since $\theta_j \in [-\lambda, \lambda]$ for $j = 1, \ldots, J$ the volume

[10] One need not employ the same explanatory variables across cases; i.e., one could use a different $\mathbf{w}_k$ for each issue area. Also, the assumption of independent Normal regressions can be relaxed by employing a seemingly unrelated regression (SUR) model. One could assume:

$$\begin{bmatrix} \alpha_k \\ \boldsymbol{\beta}_k \end{bmatrix} \sim \mathcal{N}_{D+1}(\mathbf{X}_k \boldsymbol{\xi}, \boldsymbol{\Sigma})$$

This more general model can also be estimated using standard techniques.

of S_l equals the volume of $S_{l'}$ for all $l, l' \in \{1, 2, \ldots, J!\}$.

Our approach to specifying a prior distribution on the orderings of $\boldsymbol{\theta}$ is to assume that all values of $\boldsymbol{\theta}$ that produce the same ordering are equally likely; i.e., the density of $\boldsymbol{\theta}$ is uniform within each S_l for $l = 1, \ldots, J!$. Since we believe different orderings are more likely than others *a priori*, it will generally not be the case that the prior ordinate of a point in S_l will equal the prior ordinate of a point in $S_{l'}$ for $l \neq l'$. Since the volume of S_l equals the volume of $S_{l'}$ for all $l, l' \in \{1, 2, \ldots, J!\}$, the density of a point $\boldsymbol{\theta} \in S_l$ will be proportional to the the prior probability that the ordering implied by $\boldsymbol{\theta}$ occurs. Without additional constraints, the implied density of $\boldsymbol{\theta}$ would have $J!$ parameters. Clearly, when J is greater than four, it will be very difficult to specify all $J!$ parameters directly.

We deal with the $J!$ parameters of the prior distribution for $\boldsymbol{\theta}$ by reparameterizing in terms of J parameters. More specifically, we assume that:

$$\Pr\left(\theta_{\psi(1)} < \theta_{\psi(2)} < \cdots < \theta_{\psi(N)}\right) = \prod_{i=1}^{N} \frac{\delta_{\psi(i)}}{\sum_{j=1}^{i} \delta_{\psi(j)}}, \tag{4}$$

where ψ is some permutation of $(1, 2, \ldots, J)$, and $\delta_j \in \mathbb{R}^+,\ j = 1, 2, \ldots, J$ are the new parameters governing the prior distribution of $\boldsymbol{\theta}$.[11] To evaluate this prior density (modulo a normalizing constant) at some $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, we simply look at the ranks of each element of $\boldsymbol{\theta}$ and then calculate the probability of seeing that ordering of $\boldsymbol{\theta}$ according to Equation 4. Note that it is actually quite easy to calculate the normalizing constant of this density since the the density is uniform within areas of equal volume. One can show that the normalizing constant is $\frac{1}{J!}(2\lambda)^J$; this fact will prove useful when computing the marginal likelihood.

This prior distribution can be generalized to the case of multidimensional $\boldsymbol{\theta}_j$s in a straightforward manner. If it is reasonable to assume that the ranks on a given dimension are *a priori* independent of the ranks on all other dimensions, then the joint posterior of of the multidimensional ideal points is proportional to the product of the probabilities of the ranks on each dimension. If independence of the ranks across dimensions is not reasonable *a priori*, then the prior must be

[11] This parameterization is used in a different context by Hartigan (1968).

constructed from the ranks of the JD elements of $\boldsymbol{\theta}$ in the same manner as above, where D is the dimension of the issue-space.

Eliciting Prior Beliefs Over the Orderings of Justices' Ideal Points

One of the major advantages of adopting the particular parameterization of the prior distribution over the orderings above is the relative ease of specifying an informative prior distribution over the justices' ideal points. Not only are there a relatively small number of parameters, but simple functions of these parameters coincide with quantities that judicial scholars have thought extensively about. For instance, the probability that justice i's ideal point is farthest to the right is simply $\delta_i / \sum_{j=1}^{J} \delta_j$. Further, the probability that justice i is to the right of justice j is $\delta_i/(\delta_i + \delta_j)$.

Event	Subjective Probability	Implied Probability
(Rehnquist 9th)	0.50	0.46
(Burger 9th)	0.30	0.32
(O'Connor 9th)	0.10	0.10
(Powell 9th)	0.10	0.08
(Brennan > Marshall)	0.70	0.70
(Stevens > Brennan)	0.65	0.65
(Blackmun > Stevens)	0.55	0.55
(White > Blackmun)	0.65	0.65
(Powell > White)	0.85	0.85
(O'Connor > Powell)	0.55	0.55
(Burger > O'Connor)	0.75	0.76
(Rehnquist > Burger)	0.55	0.59

Table 1: *Events, Elicited Subjective Probabilities, and the Probabilities Implied by the Value of* $\boldsymbol{\delta}$ *Used in the Analysis for the One-Dimensional Model and the Civil Liberties Dimension of the Mixture Model. The value of* $\boldsymbol{\delta}$ *used in the analysis is* (0.0696615, 0.1625898 ,0.3022852, 0.3703024, 0.6906369, 4.0059537, 4.8924510, 15.3093377, 22.4163287).

Our approach to eliciting a prior distribution for $\boldsymbol{\theta}$ from expert opinion was to ask a court scholar to furnish his prior beliefs about the probabilities of a number of events that are simple functions of $\boldsymbol{\delta}$. One approach would be to elicit subjective probabilities for J events and then solve for the J elements of $\boldsymbol{\delta}$. However, in some cases we found that such a procedure produced implicit subjective probabilities for other events that were not entirely consistent with the experts prior beliefs. The reason for this has to do with the problems that most human subjects have in thinking about probabilities, and with the fairly strong structure the parameterization discussed above puts on the distribution of $\boldsymbol{\theta}$.

Another approach (and the one adopted in what follows) is to elicit subjective probabilities for a number of events larger than J, and to then numerically find the value of $\boldsymbol{\delta}$ that implies probabilities closest (in some sense) to the elicited probabilities. More specifically, we asked a "court scholar" (in this case Martin) to provide his subjective beliefs regarding 12-13 events. With these subjective probabilities in hand, we use the BFGS algorithm to minimize an objective function which was simply the squared difference between the subjective probabilities and the actual probabilities implied by a given value of $\boldsymbol{\theta}$. This procedure yields a value of $\boldsymbol{\delta}$ that is generally fairly consistent with the subject's prior beliefs.

Event	Subjective Probability	Implied Probability
(Rehnquist 9th)	0.50	0.47
(White 9th)	0.20	0.19
(O'Connor 9th)	0.10	0.10
(Stevens 9th)	0.10	0.08
(Powell 9th)	0.10	0.06
(Blackmun $>$ Marshall)	0.60	0.60
(Burger $>$ Blackmun)	0.70	0.70
(Brennan $>$ Burger)	0.60	0.60
(Powell $>$ Brennan)	0.55	0.56
(Stevens $>$ Powell)	0.55	0.55
(O'Connor $>$ Stevens)	0.55	0.55
(White $>$ O'Connor)	0.65	0.66
(Rehnquist $>$ White)	0.70	0.72

Table 2: *Events, Elicited Subjective Probabilities, and the Probabilities Implied by the Value of* $\boldsymbol{\delta}$ *Used for the Economics Dimension of the Mixture Model. The value of* $\boldsymbol{\delta}$ *used in the analysis is* (0.3806479, 2.0644045, 3.2413899, 0.5727564, 7.7935779, 2.6103830, 4.0115578, 1.3502685, 19.7557415).

The events, subjective probabilities, the adjusted probabilities, and the associated values of $\boldsymbol{\delta}$ for the one-dimensional model and the first dimension of the mixture model are presented in Table 1. The events, subjective probabilities, adjusted probabilities, and the associated values of $\boldsymbol{\delta}$ for the second dimension of the mixture model are given in Table 2.

Results for One-Dimensional Models

So far we have presented a handful of innovative statistical models designed to alleviate the 'micro-committee' problem. But proposing models without fitting them to the data and investigating their performance contributes nothing to our knowledge of the Supreme Court. In the remainder of the paper, we present our parameter estimates from the models developed in the previous sections. After presenting our findings, we compare the performance of the models, and demonstrate how our estimates can be used to study substantive cases.

We begin by estimating one-dimensional models for the seventh Burger Court. We do so not because we believe the issue space is uni-dimensional. Since the publication of Schubert's (1965) seminal analysis, most researchers have found at a minimum a two-dimensional space: one general liberalism measure, most closely related to civil liberties, and a measure that captures economic considerations. Nonetheless, our one dimensional models serve as the baseline for comparison when we turn to the more general case.

In Table 3 we summarize the posterior density for the two one-dimensional item response models.[12] For the standard model reported in the first four columns, we assume that the ideal points θ_j and the case specific parameters α_k and β_k are drawn from independent Gaussian distribution. The estimated ideal points are the posterior mean, which is in the first column of the data. This model tells us that *a posteriori* Marshall is the most liberal justice, in fact much more liberal then Brennen, the next most liberal justice. Three pairs of justices – Stevens and Blackmun, White and Powell, and O'Connor and Burger – have ideal points very close to each other. Rehnquist is the most conservative justice, far to the right of Burger. To judge the certainty of our estimates one can look at the posterior standard deviation or the Bayesian Credible Interval (BCI). As one would expect, the model estimates ideal points best for the justices in the center of the policy space, and is less efficient for justices at the extremes. The 95% BCI gives the interval in which the parameter

[12]To alleviate the rotational invariance problem, we fix Marshall to have an ideal point below zero, which means we obtain a conservatism measure.

	Standard Model				Hierarchical Model			
	Post.	Post.	95% BCI		Post.	Post.	95% BCI	
Justice	Mean	StD.	Lower	Upper	Mean	StD.	Lower	Upper
Marshall	-2.964	0.347	-3.706	-2.341	-1.241	0.287	-1.878	-0.765
Brennan	-2.169	0.236	-2.659	-1.744	-0.902	0.206	-1.367	-0.555
Stevens	-0.451	0.094	-0.640	-0.269	-0.148	0.069	-0.297	-0.028
Blackmun	-0.241	0.091	-0.420	-0.065	-0.049	0.065	-0.184	0.073
White	0.699	0.105	0.496	0.910	0.420	0.102	0.245	0.650
Powell	0.975	0.123	0.746	1.227	0.588	0.132	0.364	0.881
O'Connor	1.486	0.162	1.186	1.826	0.789	0.168	0.501	1.151
Burger	1.726	0.187	1.380	2.106	0.940	0.199	0.599	1.380
Rehnquist	3.430	0.433	2.661	4.354	1.708	0.351	1.108	2.489
N				9				9
K				483				483
Burnin Iterations				25000				25000
Gibbs Iterations				50000				50000

Table 3: *Posterior density summary for one-dimensional item response models with standard Normal priors on the ideal points. In the standard model,* α_k *and* β_k *are assumed to come from a standard Normal distribution. The hierarchical model uses issue-specific covariates to model the distribution of these parameters (note that the hyperparameters are not reported). BCI denotes the Bayesian Credible Interval.*

lies with 95% certainty.

In addition to reporting parameter estimates and measures of our certainty about them, another important tool one can use when performing Bayesian inference is computing posterior probabilities over other important quantities that are functions of the parameters. One quantity that many scholars are interested in is the median justice on the Court. Indeed, most bargaining models of judicial behavior, not only among the justices, but between the Court and the other institutions in the separation of powers system, require a measure of the median justice. We thus record the median justice for each iteration of the Gibbs sampler and average across the iterations. In so doing, we average parameter uncertainty into this quantity of interest. We find with 98.3% certainty that White is the median justice on this dimension. There is a 1.6% chance that Powell is the median justice.

Justice	Post. Mean	Post. Std.Dev.	95% BCI Lower	Upper
Marshall	-3.440	0.324	-3.957	-2.751
Brennan	-2.436	0.290	-3.044	-1.895
Stevens	-0.468	0.099	-0.669	-0.281
Blackmun	-0.258	0.095	-0.442	-0.068
White	0.714	0.107	0.506	0.926
Powell	0.998	0.123	0.770	1.249
O'Connor	1.540	0.168	1.229	1.882
Burger	1.822	0.195	1.467	2.215
Rehnquist	3.637	0.277	2.971	3.985
N				9
K				483
Burnin Iterations				25000
Gibbs Iterations				250000

Table 4: *Posterior density summary for one-dimensional item response models with an ordinal prior on the ideal points. BCI denotes the Bayesian Credible Interval.*

As discussed earlier, due to the 'micro-committee' problem we will likely have difficulty estimating the case specific parameter. The second model presented in Table 3 contains summarizes the posterior density of a one-dimensional hierarchical model. Instead of assuming the α_k and β_k parameters are drawn from a standard Normal distribution, we model their distributions using a linear regression. We employ case-specific covariates drawn from the *United States Supreme Court Database*. We construct seven dichotomous variables that denote whether a case relates to criminal procedure, civil rights, the First Amendment, due process, economics, judicial power, and federalism. These variables, along with a constant, forms $Q \times 1$ vector $\mathbf{w}_k$ for each case. We use the sample covariates for the other hierarchical model, and the mixture model with covariates, reported below.

To save space, we only report the θ_j coefficients in Table 3. While the estimated ideal points are on a different scale, it is clear that the model tells the same substantive story about the ideal points of the justices. There are a handful of interesting findings when looking at the hyperparameters.

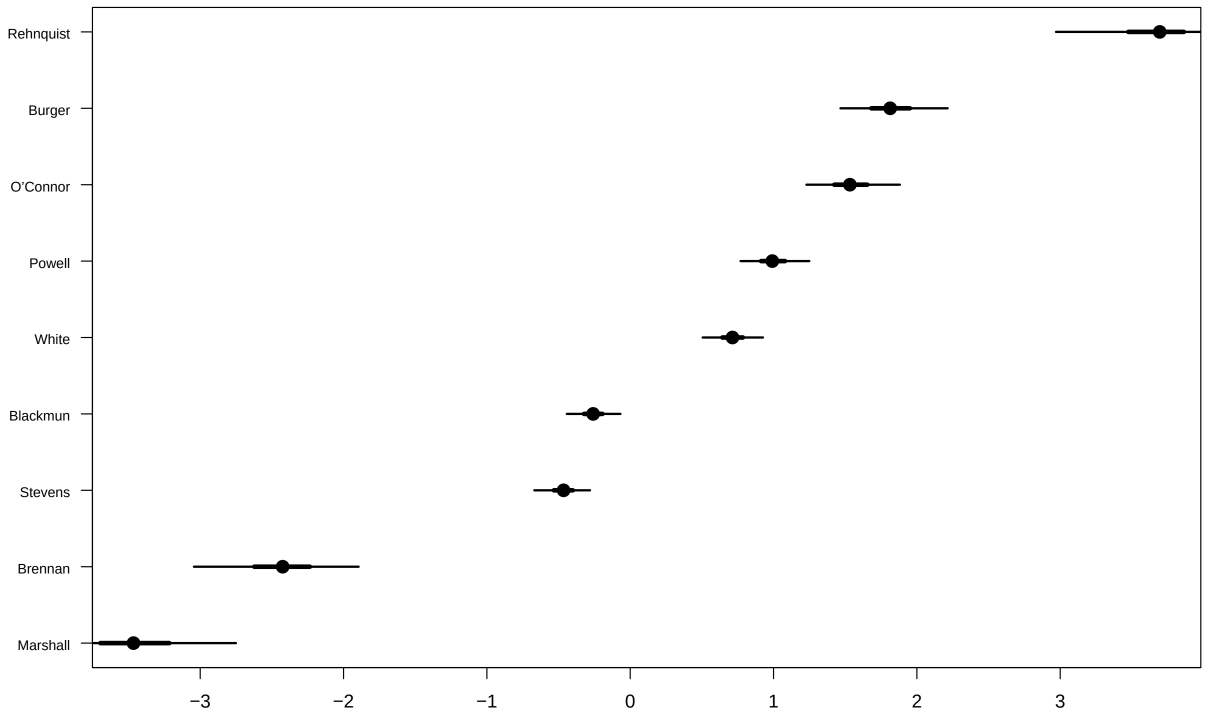

Figure 1: *Lineplot of posterior densities of justices' ideal points for the one dimensional model with an ordinal prior. The dot represents the posterior median, the large band represents the central 50% of the distribution, and the thin line represents the central 95% of the distribution.*

First, none of the hyperparameters achieve statistical significance in the α_k equation, but all of them do in the β_k equation. This implies that the covariates do little in explaining the former parameter, but do contribute to our knowledge of the latter. Moreover, for 51.7% of the case-specific parameters we get shrinkage; i.e., we get posterior densities with less variance.

The final one-dimensional model we estimate is one with a strong prior on the ordering of the justices. We employ the prior described in Table 1. These results are presented in Table 4. When looking at the ideal points we again see the same structure, with Marshall on the far left, Rehnquist on the far right, and the same ordering. The improvement with this model is in terms of precision. Not only are the posterior standard deviations smaller, but the BCIs are narrower for each justice (after taking into account differences of scale). This is a good example of how using

	1A	1B	2	3A	3B	4A	4B
Marshall	86.6	85.7	87.6	89.6	88.7	90.0	89.6
Brennan	83.3	81.9	84.0	84.8	84.2	85.7	85.8
Stevens	61.1	59.6	61.2	80.2	68.4	69.3	69.4
Blackmun	66.3	64.2	66.4	69.3	69.2	68.8	68.5
White	68.6	67.1	68.7	70.4	69.6	70.2	70.1
Powell	72.7	72.1	72.7	73.4	74.3	74.8	74.2
O'Connor	77.8	76.5	77.8	78.0	77.0	78.3	78.1
Burger	79.8	79.3	79.9	81.1	83.0	83.0	82.6
Rehnquist	87.5	87.4	87.6	87.3	87.0	86.4	86.5
Overall	76.0	74.8	76.2	79.3	77.9	78.5	78.3

Table 5: *Percent votes correctly predicted for each justice for all seven models. Models 1A and 1B are the standard and hierarchical one-dimensional models. Model 2 is the one-dimensional model with an ordinal prior. Models 3A and 3B are the two-dimensional models with Normal and hierarchical priors. Models 4A and 4B are the mixture models without and with covariates, respectively.*

prior information can greatly improve the estimates one gets from a statistical model. We again compute the posterior probability distribution over the median justice, and find that White is the medial justice with 99.4% probability, and 0.6% probability that Powell is the median. In Figure 1 we summarize this posterior density with a line plot. While there is overlap in the credible intervals for four pairs of justices, there is substantial differentiation across the justices.

While looking at coefficient estimates across models is informative, what are really interested in is how well each model predicts votes on the Court. In Table 5, we present the percent correctly predicted for each justice as well as the overall percentage for each model presented in this paper. Models 1A and 1B in this table are the standard model with Normal and hierarchical priors. Surprisingly, for all justices except for White (the median justice), and for the model overall, the model with Normal priors does a better job predictively than the hierarchical model. There seems to exist a tradeoff; we get better estimates of the case-specific parameters, but do a poorer job of prediction. We suspect this is because our covariates are only partially informative, and because our sample size is relatively small. Model 2 represents the one-dimensional model with ordinal priors. For all justices and overall, this model outperforms the model with Normal priors.

Results for Two-Dimensional Models

In Table 6 we summarize the posterior densities for two dimensional models with standard Normal and hierarchical priors on the case-specific parameters.[13] For the standard model, we see first dimension estimates similar to those for the one-dimensional models. The justices retain the same ordering. As expected, the posterior standard deviations are greater for the two-dimensional model than the one-dimensional model (after taking scale into account). The second dimension estimates are more interesting. Indeed we see an ordering of the justices similar to the first dimension *except* for Stevens, who is much more conservative on this dimension. By inspecting these scales, it is clear that the first dimension is a general civil liberties dimension, and the second dimension measure economic conservatism.

The second model in Table 6 is a hierarchical model where we use our case-specific dummy variables to model te distributions of the α_k and $\boldsymbol{\beta}_k$ parameters. The results from this model are quite poor. The estimated $\boldsymbol{\theta}_j$ have large Bayesian credible intervals. Moreover, inspecting the hyperparameters (not reported in the table) shows that the covariates perform poorly in explaining the case-specific parameters. Close inspection of the posterior density sample suggests that convergence is an issue. By including case-specific information, and thus estimating an additional $3 \times (Q + 1)$ additional parameters, it is likely that the Markov chain has not reached steady state. While this model does reasonably well predictively, these results cannot be trusted for the reasons stated above.

The final results we present are for two-dimensional mixture models with ordinal priors. These results, in Table 7 are quite interesting. The first four columns of the table include results from the mixture model without covariates. We obtain first dimension results similar to those we have seen before. The second dimension scores, however, depart markedly from the two-dimensional model with Normal priors. Marshall is the most liberal justice on this dimension, followed next by

[13]To fix the rotational invariance problem, we put three constraints on the sampler. We force Marshall to have a negative score on both dimensions, and constrain Stevens to have dimension two score greater than his dimension one score.

Blackmun, Burger, and Powell. All justices have ideal points that fall below zero except for Stevens, who is the most conservative on this dimension. We again suspect this is an economic dimension. For the mixture model without covariates, there is a 99.3% posterior probability that White is the median justice, and a 48.9% chance that he is the median justice on the second dimension. There is a 26.6% chance that Brennan is the median justice on the second dimension.

In the second model we use our case-specific covariates to model the probability that each case falls on each of the two dimensions. The vast majority of our cases fall on the first dimension. Indeed, the model indicates that only 7.6% of the cases have a greater than 80% chance of falling on the second dimension; the first dimension dominates the politics of the seventh Burger Court. The logit coefficients for the mixture model are most interesting. Criminal procedure and First Amendment cases are significantly more likely to fall on the first dimension; economics and judicial power cases are significantly more likely to fall on the second. The posterior probabilities of the median justice are similar to the model without covariates. White is most likely to be the median justice on the first dimension (99.2%). He also has a 35.7% chance of being the median justice on the second dimension. Brennan has a 27.7% chance, and Powell has a 23.4% chance.

One can compare the predictive power of the models by referring back to Table 5. One notices that the two-dimensional models perform better than the one-dimensional models, but only by a small percentage. This is not surprising, as only a small fraction of the cases fall on the second dimension. The two-dimensional model with independent Normal priors does the best predictively. Nonetheless, the mixture model with covariates does a superior job of estimating the case-specific parameters. Again there is a trade-off; by estimating more parameters we do a poorer job of prediction, while extracting more substantive information about each case. Additionally, the theoretical assumption of equal issue salience required for the two-dimensional with Normal priors is a difficult one to swallow. For this reason, we use the the results from the mixture model with covariates to illustrate the use of the case-specific parameters to study individual cases.

	Standard Model				Hierarchical Model			
	Post.	Post.	95% BCI		Post.	Post.	95% BCI	
Justice	Mean	StD.	Lower	Upper	Mean	StD.	Lower	Upper
Dimension One								
Marshall	-2.659	1.019	-4.318	-0.501	-0.499	0.243	-1.069	-0.145
Brennan	-1.828	0.608	-2.754	-0.497	-0.344	0.221	-0.856	-0.021
Stevens	-1.397	0.607	-2.661	-0.297	-0.544	0.451	-1.673	-0.062
Blackmun	-0.082	0.237	-0.486	0.419	0.310	0.183	-0.013	0.665
White	0.747	0.153	0.452	1.044	0.583	0.191	0.242	0.987
Powell	0.892	0.178	0.507	1.212	0.704	0.249	0.260	1.225
O'Connor	1.227	0.293	0.559	1.692	0.694	0.216	0.315	1.169
Burger	1.623	0.294	0.972	2.159	1.186	0.388	0.487	2.018
Rehnquist	2.733	0.714	1.207	3.957	1.116	0.300	0.620	1.806
Dimension Two								
Marshall	-2.218	1.118	-4.194	-0.196	-1.468	0.422	-2.341	-0.625
Brennan	-1.288	0.751	-2.586	0.002	-1.013	0.309	-1.654	-0.458
Stevens	1.101	0.762	-0.457	2.433	0.267	0.312	-0.096	1.058
Blackmun	-0.465	0.202	-0.805	0.023	-0.443	0.206	-0.868	-0.123
White	-0.038	0.354	-0.678	0.595	-0.050	0.154	-0.346	0.234
Powell	0.198	0.381	-0.513	0.855	0.062	0.196	-0.287	0.467
O'Connor	0.537	0.503	-0.369	1.410	0.277	0.215	-0.096	0.712
Burger	0.363	0.672	-0.831	1.517	0.185	0.253	-0.277	0.651
Rehnquist	1.293	1.076	-0.572	3.152	0.891	0.330	0.371	1.583
N				9				9
K				483				483
Burnin Iterations				5000				5000
Gibbs Iterations				50000				50000

Table 6: *Posterior density summary for two-dimensional item response models with independent standard Normal priors on the ideal points. In the standard model,* α_k *and* $\boldsymbol{\beta}_k$ *are assumed to come from independent standard Normal distributions. The hierarchical model uses issue-specific covariates to model the distribution of these parameters (note that the hyperparameters are not reported). BCI denotes the Bayesian Credible Interval.*

	Mixture Model				Mixture Model with Covariates			
	Post.	Post.	95% BCI		Post.	Post.	95% BCI	
Parameter	Mean	StD.	Lower	Upper	Mean	StD.	Lower	Upper
Dimension One								
Marshall	-3.818	0.162	-3.995	-3.394	-3.830	0.150	-3.995	-3.445
Brennan	-2.906	0.356	-3.629	-2.242	-2.919	0.364	-3.732	-2.266
Stevens	-0.519	0.102	-0.724	-0.325	-0.515	0.104	-0.725	-0.318
Blackmun	-0.247	0.097	-0.440	-0.056	-0.251	0.097	-0.440	-0.060
White	0.784	0.118	0.564	1.021	0.796	0.120	0.565	1.037
Powell	1.106	0.136	0.850	1.381	1.098	0.146	0.823	1.399
O'Connor	1.683	0.178	1.355	2.051	1.670	0.183	1.335	2.052
Burger	2.133	0.235	1.722	2.624	2.131	0.243	1.653	2.616
Rehnquist	3.771	0.188	3.289	3.992	3.835	0.147	3.461	3.995
Dimension Two								
Marshall	-3.363	0.500	-3.979	-2.186	-2.979	0.687	-3.954	-1.464
Brennan	-1.448	0.490	-2.634	-0.672	-1.613	0.498	-2.687	-0.748
Stevens	3.460	0.446	2.396	3.985	3.419	0.491	2.144	3.981
Blackmun	-3.294	0.542	-3.973	-2.014	-2.758	0.662	-3.916	-1.485
White	-1.844	0.562	-3.071	-0.904	-1.666	0.533	-2.895	-0.801
Powell	-2.532	0.673	-3.757	-1.286	-1.916	0.599	-3.211	-0.878
O'Connor	-0.843	0.362	-1.637	-0.224	-0.846	0.374	-1.681	-0.199
Burger	-2.859	0.662	-3.916	-1.452	-2.483	0.717	-3.848	-1.172
Rehnquist	-0.407	0.288	-1.002	0.137	-0.517	0.325	-1.215	0.063
Mixture Logit Coefficients								
γ-Constant					-1.461	0.483	-2.489	-0.596
γ-Criminal Procedure					-2.926	1.515	-6.966	-0.845
γ-Civil Rights					-1.792	1.282	-5.220	0.013
γ-First Amendment					-2.619	1.846	-7.123	0.012
γ-Due Process					-1.960	1.720	-6.070	0.561
γ-Economics					1.023	0.580	-0.058	2.200
γ-Judicial Power					1.025	0.628	-0.187	2.306
γ-Federalism					0.760	0.968	-1.249	2.541
N				9				9
K				483				483
Burnin Iterations				25000				25000
Gibbs Iterations				250000				250000

Table 7: *Posterior density summary for mixture item response models with ordinal priors on the ideal points. The model with covariates uses issue-specific covariates to model the probability that each case falls on the second dimension. The reported* γ *coefficients are logit coefficients that predict this probability. BCI denotes the Bayesian Credible Interval.*

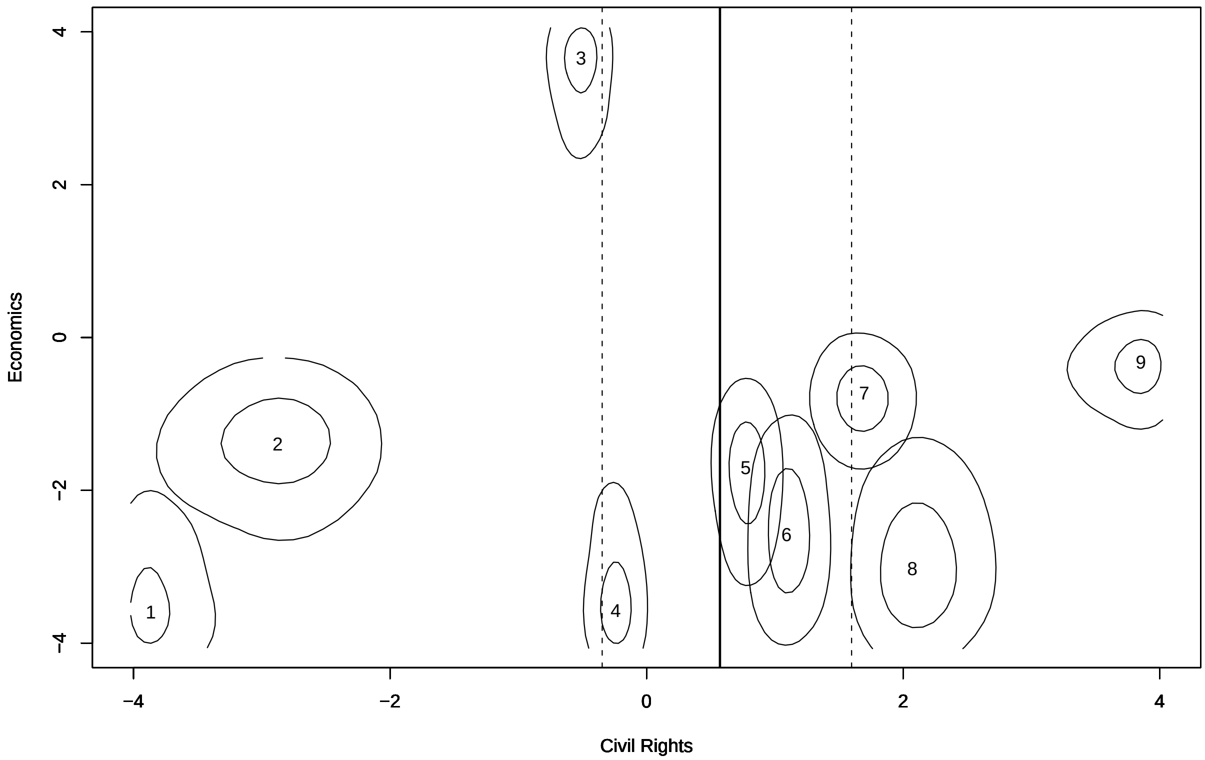

Figure 2: *Spatial representation of Garcia v. San Antonio Metropolitan Transit Authority (469 US 528). The justices' ideal points are estimated using the mixture model with an ordinal prior. The posterior probability that this case falls along the first dimension is over 99%, so we only report the first dimension cutpoint. For each justice, the confidence ellipses are the 50% and 95% highest density regions. The solid vertical line represents the estimated posterior median cutpoint for this case, and the dashed lines denote its 90% credible interval. The justices are ordered: Marshall, Brennen, Stevens, Blackmun, White, Powell, O'Connor, Burger, and Rehnquist.*

In Figure 2 we present our estimated spatial representation for *Garcia v. San Antonio Metropolitan Transit Authority* (469 US 528). Our model predicts that this case falls on the first dimension with greater than 99% probability. The majority in this case decided that provisions of the Fair Labor Standards Act apply to sub-national governments. The majority opinion was written by Blackmun, who was joined by Brennan, White, Marshall, and Stevens. In the figure, each justice is represented by a number. The small shape surrounding the number denotes the 50% confidence region of the measure, and the large shape denotes the 95% confidence region. For all justices but Brennan and Rehnquist, these regions are much longer on dimension two than dimension one. This

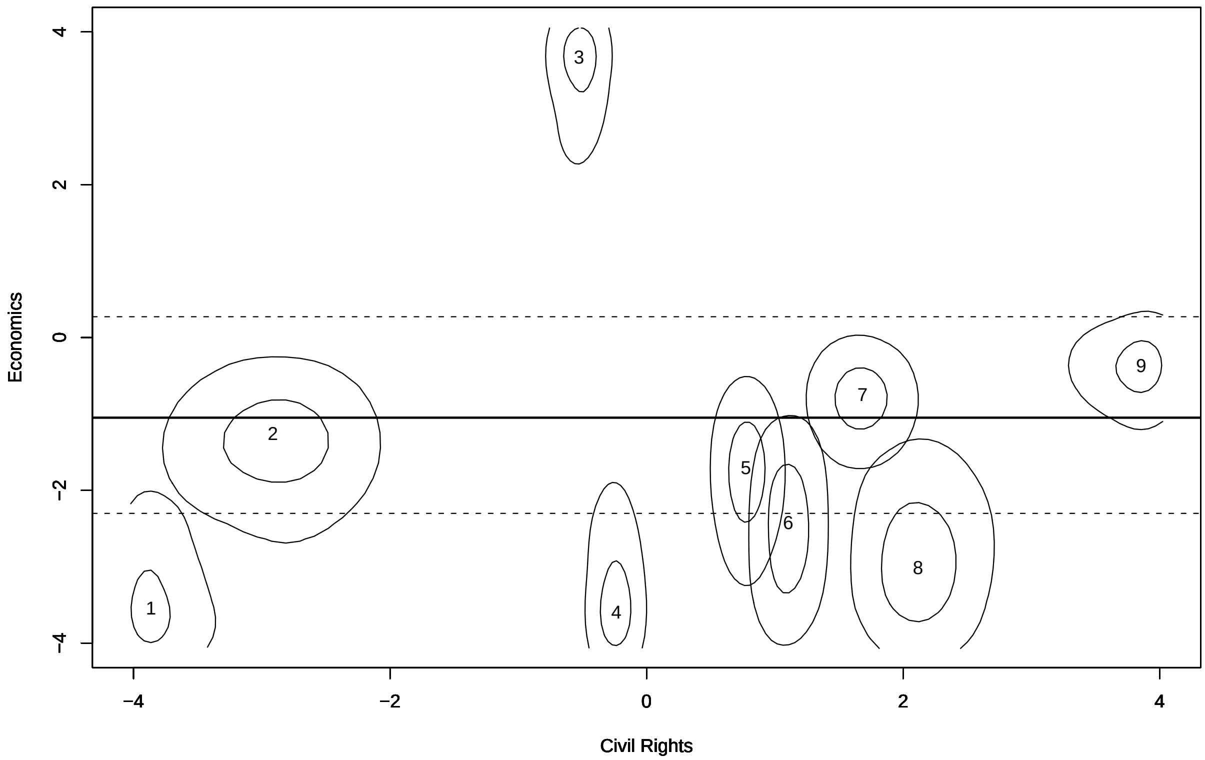

Figure 3: *Spatial representation of Dixson v. United States (465 US 482). The justices' ideal points are estimated using the mixture model with covariates and an ordinal prior. The posterior probability that this case falls along the second dimension is over 86%, so we only report the second dimension cutpoint. For each justice, the confidence ellipses are the 50% and 95% highest density regions. The solid horizontal line represents the estimated posterior median cutpoint for this case, and the dashed lines denote its 90% credible interval. The justices are ordered: Marshall, Brennen, Stevens, Blackmun, White, Powell, O'Connor, Burger, and Rehnquist.*

means that the model is doing a better job of estimating first dimension ideal points than second dimension points. The vertical black line in the center of the figure denotes the estimated cutpoint for this case. The model predicts that all justices on the left of the cutpoint will vote to reverse the lower court, and those on the right would affirm. The model correctly predicts eight of the nine justices. The posterior median ideal point for White, the median justice, falls to the right of the cutting line, so we would expect White to vote to affirm. Yet, the 99% confidence region intersects the cutpoint. Since White is the pivotal justice, it would be interesting to examine the memoranda from this case to see whether White was able to gain concessions from the more liberal justices.

We consider a case on the economics dimension in Figure 3: *Dixson v. United States* (465 US 482). This case, which related to whether officers of private non-profit corporations can be subject to prosecution under federal bribery statutes, is coded as an economic activity case by Spaeth (1999). We estimate that this case falls on the second dimension with 86% posterior probability. Marshall wrote the opinion of the Court, and was joined by Burger, White, Blackmun, and Powell in affirming the decision of the lower Court. O'Connor wrote the dissent, joined by Stevens, Rehnquist, and Brennan. Because this case falls on the second dimension, the cutpoint is a horizontal line. We see a great deal of overlap in the preference estimates for the justices, likely due to the small number of cases decided on this dimension. Based on the posterior medians, the model predicts all votes correctly in this case except for Brennan. Yet, there is some posterior probability that he would vote to reverse in this case.

Certainly these figures are not the definitive explanations for voting on either of these cases. Rather, we use them to illustrate how one can use the estimated case-specific parameters to understand voting on individual cases. One thing that is clear, based on this analysis, is that our estimates of the second dimension ideal points are quite imprecise. This is due to the small number of cases that seem to be decided on this dimension. As we discuss below, one solution to this problem is extending the time period of our study, thus adding more data points to the analysis.

Conclusions and Future Research

In this paper we have posited a simple theoretical model of Supreme Court decision making that is consistent with both attitudinal and strategic explanations. We have used this theoretical model to derive a statistical model of the justices' decisions at the individual justice-case level. This statistical model is not only consistent with the theoretical model, but it makes exactly the same structural assumptions. As a result, if the theoretical model is correct, the statistical model allows us to gauge the extent to which justices' policy preferences and case attributes determine decisions on the merits.

Unlike most previous work on preference estimation in political science, our goal is not to produce and supply pre-packaged measures of the ideal points of political actors.[14] Instead, we hope to provide an initial attempt at a framework that can be used to answer a wide range of questions in which the decisions of Supreme Court justices (or other political decision makers for that matter) can be viewed as the ultimate dependent variable. The models that we have presented in this paper are all fairly simple to the extent that the only variables on the right-hand side are the latent ideal points and case parameters. Nonetheless, it is a trivial matter to include other covariates directly, or to parameterize the prior distribution of the ideal points or case parameters in terms of additional variables such as social background variables or case facts.[15]

In addition, this paper has provided modeling strategies that offer promise at mediating the statistical problems associated with estimating latent preferences with a small number of decision makers. One strategy that we believe can be greatly improved upon is the idea of using a hierarchical prior for the case parameters to borrow strength across similar cases and consequently produce more accurate estimates of the case parameters as well as the latent ideal points. Another innovation is the use of a mixture of item response models to account for the potential multidimensionality of the underlying issue space. Such a modeling strategy can be motivated naturally from a structure induced equilibrium perspective, and allows us to assess the evidence in favor of almost any number of dimensions – something not feasible using a standard multidimensional item response model with nine decision makers. The key to successfully implementing the mixture model is the specification of a prior distribution over the ordering of the justices' ideal points. This prior is informative enough to prevent label-switching, and is relatively easy to elicit from court scholars.

Substantively, our findings indicate that there is evidence of at least two underlying dimensions to the preferences of justices serving on the seventh natural Burger court. Our results suggest that these two dimensions are best thought of as a civil liberties dimension and an economics dimension. We find that (depending on the modeling assumptions) somewhere between 80% and 90% of the

[14]Although, if the spatial model of judicial choice we posit is true, one could reliably use these measures in other contexts.

[15]For instance, see Clinton et al. (2000) in the context of congressional roll call voting.

483 cases under study were decided on the civil liberties dimension. Predictively, our models do well – correctly classifying between 75% and 80% percent of the decisions. This is a substantial improvement over the observed marginal frequency of $y = 0$, which was 0.54. Interestingly, the posterior modal ordering of the justices on the civil liberties dimension (Marshall, Brennan, Stevens, Blackmun, White, Powell, O'Connor, Burger, Rehnquist) is similar, but not entirely consistent with, the ordering implied by Segal/Cover scores. In particular, Blackmun appeared much more conservative at confirmation, while White and O'Connor appeared much more liberal.

This paper is very much a first step at what we hope will be a fruitful approach to the analysis of Supreme Court decision-making and micro-committee decision-making more generally. In future research, we plan to explore a number of obvious (and some perhaps not so obvious) extensions to the models discussed above. For instance, there is clearly much room for improving the hierarchical prior on the case parameters. By including better covariates, or perhaps reparameterizing the case parameters, we hope to substantially improve the fit of the models in question. Similarly, it seems natural to use the information regarding the opinion writer of each decision in the hierarchical prior for the case parameters. One approach would be to assume that the opinion writer is choosing the position of the policy alternative to maximize his or her expected utility. The position that maximizes the opinion writer's expected utility could then be treated as the mean of the prior distribution for the policy position under reversal of the lower court ruling. A related approach would posit a game theoretic model of opinion assignment, opinion writing, and voting on the merits. By invoking a solution concept such as quantal response equilibrium (McKelvey and Palfrey 1995, 1996, 1998) it would be possible to fit the implied statistical model (Signorino 1999). Of course, such approaches would require reparameterizing the case parameters, but is in principle possible. Such approaches are similar to that suggested by Londregan (2000a,b).

Another extension that we are planning is to extend the model to handle longitudinal data. Not only does this provide us with more data on both cases and justices, but it also allows us to compare the preferences of justices who *never* served together. For example, while Brandeis did not serve with Brennan; Brandeis served with Butler, who served with Reed, who served with Clark, who served with Brennan. Such overlapping generations of justices do, in principle, allow us to measure Brandeis' preferences on the same scale as Brennan's preferences [for a related example see Berry et al. (1999)]. Another natural extension that longitudinal data would permit is to allow justice ideal points to change over time, something suggested in the judicial literature by Epstein et al. (1998), and in the congressional literature by Poole and Rosenthal (1997) with their D-NOMINATE scores. We are also interested in developing improved methods to elicit and pool the prior beliefs of court scholars. Of critical importance here are questions regarding the robustness of our results as well as the psychological aspects of accurately eliciting expert opinions.

Finally, we plan to use Bayes factors (Kass and Raftery 1995) to compare models and to gauge the dimensionality of the issue space. At this point, we have successfully implemented the methods of Chib (1995) and Chib and Jeliazkov (2001) for a handful of the models under study. Such an approach offers a principled approach to the determination of the number of underlying issue dimensions – something that is generally lacking in psychometric literature [but see Oh and Raftery (2000)].

Appendix A. MCMC Estimation for the Basic Item Response Model

We are interested in the posterior density of the standard item response model:

$$\begin{aligned} f(\mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta}|\mathbf{Y}) &\propto f(\mathbf{Y}|\mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})\pi(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \\ &\propto L(\mathbf{Y}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) \prod_{j=1}^{J} \pi(\boldsymbol{\theta}_j) \prod_{k=1}^{K} \pi(\alpha_k, \boldsymbol{\beta}_k) \\ &\propto \prod_{j=1}^{J} \prod_{k=1}^{K} \left\{ \Phi(\alpha_k + \boldsymbol{\beta}_k' \boldsymbol{\theta}_j)^{y_{j,k}} \left[1 - \Phi(\alpha_k + \boldsymbol{\beta}_k' \boldsymbol{\theta}_j)\right]^{1-y_{j,k}} \right\} \prod_{j=1}^{J} \pi(\boldsymbol{\theta}_j) \prod_{k=1}^{K} \pi(\boldsymbol{\eta}_k) \end{aligned}$$

Note that this does not easily simplify to a common distribution. We assume the default Normal priors discussed in the text. Our approach is to simulate a large number of values from this distribution. To do so, we employ the Gibbs sampling algorithm.

The Gibbs sampling algorithm for this item response model contains three conditional distributions from which we iteratively sample. All of them take standard forms. To aid in exposition, we need to define three additional quantities. First, let $\boldsymbol{\theta}_j^* = (1 \ \boldsymbol{\theta}_j')'$ denote a $(D+1) \times 1$ column vector that contains a constant followed by the ideal point estimate. Second, let $\boldsymbol{\eta}_k = (\alpha_k \ \ \boldsymbol{\beta}_k')'$. One should note that $\boldsymbol{\eta}_k' \boldsymbol{\theta}_j^* = \alpha_k + \boldsymbol{\beta}_k' \boldsymbol{\theta}_j$. Third, we define $\boldsymbol{\theta}^*$ to be the $J \times (D+1)$ matrix formed by stacking these elements for all j.

The full conditional distributions are:

1. $f(z_{j,k}|\mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\eta})$. For all justices j on cases k we simulate from the following distribution:

$$f(z_{j,k}|y_{j,k}, \boldsymbol{\theta}_j, \alpha_k, \boldsymbol{\beta}_k) = \begin{cases} \mathcal{N}_{[0,\infty)}(\alpha_k + \boldsymbol{\beta}_k' \boldsymbol{\theta}_j, 1) & \text{if} \quad y_{j,k} = 1 \\ \mathcal{N}_{(-\infty,0]}(\alpha_k + \boldsymbol{\beta}_k' \boldsymbol{\theta}_j, 1) & \text{if} \quad y_{j,k} = 0 \\ \mathcal{N}(\alpha_k + \boldsymbol{\beta}_k' \boldsymbol{\theta}_j, 1) & \text{if} \quad y_{j,k} \text{ is unobserved} \end{cases}$$

   Where $\mathcal{N}_{[a,b]}$ denotes the Gaussian distribution truncated on the interval $[a, b]$. This is the standard set-up for estimating a probit model from Albert and Chib (1993). This distribution tells us that we simulate from truncated Normal distributions if the vote is observed, and from an untruncated Normal if the vote is not. In performing this data augmentation, the missing data problem is ameliorated. For this step, one must loop over all justices and all cases.

2. $f(\boldsymbol{\eta}|\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta})$. For all cases k we simulate the case-specific parameters $\boldsymbol{\eta}_k$ from a $(D+1)$-variate Normal distribution:

$$f(\boldsymbol{\eta}_k|\mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}) = \mathcal{N}_{D+1}(\mathbf{e}, \mathbf{E})$$

   Where $\mathbf{e} = \mathbf{E}\left[\boldsymbol{\theta}^{*'}\mathbf{z}_{\cdot,k} + \mathbf{B}_0^{-1}\mathbf{b}_0\right]$ and $\mathbf{E} = \left[\boldsymbol{\theta}^{*'}\boldsymbol{\theta}^* + \mathbf{B}_0^{-1}\right]^{-1}$. Note that $\mathbf{z}_{\cdot,k}$ is the $(J \times 1)$ column vector of latent utilities for all justices on issue k. For this step one must loop over all cases.

3. $f(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{Z}, \boldsymbol{\eta})$. For all justices j we simulate their ideal point from a D-variate Normal distribution:

$$f(\boldsymbol{\theta}_j|\mathbf{Y}, \mathbf{Z}, \boldsymbol{\eta}) = \mathcal{N}_D(\mathbf{t}, \mathbf{T})$$

   Where $\mathbf{t} = \mathbf{T}\left[\boldsymbol{\beta}'(\boldsymbol{\alpha} - \mathbf{z}_{j,\cdot}')\right]$ and $\mathbf{T} = \left[\boldsymbol{\beta}'\boldsymbol{\beta} + \mathbf{j}_D\right]^{-1}$. Note that $\mathbf{z}_{j,\cdot}$ is a $(1 \times K)$ row vector of latent utilities of member j across all cases.

One iterates a large number of times from these conditional distributions, using draws from the previous conditionals as updating values. This is just a generalization of the Albert (1992) and Johnson and Albert (1999) algorithms to a D-dimensional issue space. Patz and Junker (1999) develop a similar algorithm for a one-dimensional model with a logit link function. Clinton et al. (2000) use a similar algorithm to simulate from the posterior density of a one-dimensional model, fit to congressional roll call voting data.

For the hierarchical model, we modify this algorithm in two ways. First, we add an additional step to the sampler to estimate the $\boldsymbol{\xi}$ and $\boldsymbol{\sigma}^2$ coefficients. Using conjugate Normal priors, the full conditional distributions of the $\boldsymbol{\xi}$ parameters are all univariate Normal. Assuming conjugate inverse-Gamma priors, the full conditional distributions of the $\boldsymbol{\sigma}^2$ parameters are all inverse-Gamma. They take the common form of a linear regression model. The second modification is requires replacing the priors $\mathbf{b}_0$ and $\mathbf{B}_0$ in Step 2 above with the estimated hyperparameters.

Appendix B. MCMC Estimation for the Item Response Model with Ordinal Prior

The sampling for this model is very similar to that discussed in Appendix A. The only additional complication arises because $p(\boldsymbol{\theta})$ is no longer conjugate to the likelihood. This makes it difficult to write down the full conditional distribution of $\boldsymbol{\theta}$ in closed form and to employ a Gibbs update.

Instead, we employ the Metropolis-Hastings algorithm with a random walk proposal. More specifically, at the $g+1$th scan of the MCMC sampling, we generate a candidate value of $\boldsymbol{\theta}$ denoted $\boldsymbol{\theta}_{can}$ from $t_\nu(\boldsymbol{\theta}^{(g)}, \tau^2\mathbf{I})$ and set $\boldsymbol{\theta}^{(g+1)}$ equal to $\boldsymbol{\theta}_{can}$ with probability:

$$\alpha(\boldsymbol{\theta}^{(g)}, \boldsymbol{\theta}_{can}|\mathbf{Y}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{Z}) = \min\left\{1, \; \frac{\prod_{k=1}^K \prod_{j=1}^J \phi(\alpha_k + \beta_k' \theta_{(can)j} - z_{jk}) p(\boldsymbol{\theta}_{can})}{\prod_{k=1}^K \prod_{j=1}^J \phi(\alpha_k + \beta_k' \theta_j^{(g)} - z_{jk}) p(\boldsymbol{\theta}^{(g)})}\right\}$$

and set $\boldsymbol{\theta}^{(g+1)}$ equal to $\boldsymbol{\theta}^{(g)}$ otherwise, where $t_\nu(\boldsymbol{\mu}, \mathbf{T})$ is the multivariate t distribution with mean $\boldsymbol{\mu}$, scale matrix $\mathbf{T}$, and ν degrees of freedom; and $\phi(\cdot)$ is the standard normal pdf. We choose ν and τ^2 to achieve an acceptance rate of approximately 25%. This Metropolis step replaces Step 2 in the algorithm presented in Appendix A; the remainder of the updates remain the same.

Appendix C. MCMC Estimation for the Mixture Item Response Model with Ordinal Prior

The MCMC sampling for the mixture item response model with and ordinal prior is greatly simplified by the addition of a second block of latent data. Note that if we knew which dimension on which each case was being decided, the mixture model would reduce to a simple item response model. Let the $\zeta_k \in \{0, 1\}$ denote an indicator of whether the kth case is being decided on the 1st issue dimension ($\zeta_k = 0$) or the second issue dimension ($\zeta_k = 1$).

With the inclusion of these auxiliary variables, one scan of an MCMC sampling scheme for this model is given by:

1. Sample from $[\mathbf{Z}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\rho}, \boldsymbol{\zeta}, \mathbf{Y}]$
2. Sample from $[\boldsymbol{\zeta}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\rho}, \mathbf{Y}, \mathbf{Z}]$
3. Sample from $[\boldsymbol{\rho}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{Y}, \mathbf{Z}, \boldsymbol{\zeta}]$
4. Sample from $[\boldsymbol{\theta}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{Y}, \mathbf{Z}, \boldsymbol{\zeta}, \boldsymbol{\rho}]$
5. Sample from $[(\boldsymbol{\alpha}, \boldsymbol{\beta})|\mathbf{Y}, \mathbf{Z}, \boldsymbol{\zeta}, \boldsymbol{\rho}, \boldsymbol{\theta}]$.

The distribution of z_{jk} given the other model parameters and latent data is given by:

$$[z_{jk}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\rho}, \boldsymbol{\zeta}, \mathbf{Y}] = \begin{cases} \mathcal{N}(\mu_{jk0}, 1)I(-\infty, 0) & \text{if } \zeta_j = 0 \text{ and } y_{jk} = 0 \\ \mathcal{N}(\mu_{jk1}, 1)I(-\infty, 0) & \text{if } \zeta_j = 1 \text{ and } y_{jk} = 0 \\ \mathcal{N}(\mu_{jk0}, 1)I(0, \infty) & \text{if } \zeta_j = 0 \text{ and } y_{jk} = 1 \\ \mathcal{N}(\mu_{jk1}, 1)I(0, \infty) & \text{if } \zeta_j = 1 \text{ and } y_{jk} = 1 \\ \mathcal{N}(\mu_{jk0}, 1) & \text{if } \zeta_j = 0 \text{ and } y_{jk} \text{ is missing} \\ \mathcal{N}(\mu_{jk1}, 1) & \text{if } \zeta_j = 1 \text{ and } y_{jk} \text{ is missing.} \end{cases}$$

where $\mu_{jkd} = \alpha_k + \beta_k' \theta_{jd}$, and $\mathcal{N}(\mu, \sigma^2)I(a, b)$ is a normal distribution with mean μ and variance σ^2 truncated to the interval $[a, b]$.

The distribution of ζ_k given the other model parameters and latent data is binomial (in the case of more than two mixture components it is multinomial) and is given by:

$$[\zeta_k|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\rho}, \mathbf{Y}, \mathbf{Z}] = \mathcal{B}inomial(1, p_k)$$

where

$$p_k = \frac{\rho_2 \prod_{j=1}^J \phi(\mu_{jk1} - z_{jk})}{\rho_1 \prod_{j=1}^J \phi(\mu_{jk0} - z_{jk}) + \rho_2 \prod_{j=1}^J \phi(\mu_{jk1} - z_{jk})},$$

$\phi(\cdot)$ is the standard normal pdf, and μ_{jkd} is as given above.

The distribution of ρ given the other model parameters and latent data is Beta since we have assumed a conjugate Beta prior with parameters $a = 1$ and $b = 1$.[16] More specifically, in the case of just two mixture components the full conditional for ρ_2 is $\mathcal{B}eta\left(a + (K - \sum_{k=1}^K \zeta_k),\ \ b + \sum_{k=1}^K \zeta_k\right)$. It follows that $\rho_1 = 1 - \rho_2$. The full conditionals for $\boldsymbol{\theta}$ and $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ are almost identical to those of the basic item response model except that now the only cases entering the update for θ_{jd} are those in which $\zeta_k = (d - 1)$. Similarly, only $\boldsymbol{\theta}_{\cdot d}$ enters the update for (α_k, β_k) where $d = (\zeta_k + 1)$.

[16] In the more general case of more than two mixture components, the conjugate prior for ρ is Dirichlet. It follows that the full conditional for ρ would also be Dirichlet.

Appendix D. MCMC Estimation for the Mixture Item Response Model with Ordinal Prior

The MCMC sampling for this model is almost identical to that discussed in Appendix C. The only major difference is that the ρ_{k1} and ρ_{k2} are parameterized as:

$$\rho_{k2} = \frac{\exp(\mathbf{w}_k'\gamma)}{1+\exp(\mathbf{w}_k'\gamma)} \qquad \text{and} \qquad \rho_{k1} = 1 - \rho_{k2}$$

and consequently are deterministic functions of $\boldsymbol{\gamma}$. Thus inference for the mixture weights reverts to inference for $\boldsymbol{\gamma}$. We adopt a $\mathcal{N}(\mathbf{0}, 10\mathbf{I})$ prior for $\boldsymbol{\gamma}$ and employ a random walk Metropolis update.

More specifically, at the $g+1$th scan of the MCMC sampling, we generate a candidate value of $\boldsymbol{\gamma}$ denoted $\boldsymbol{\gamma}_{can}$ from $t_\nu(\boldsymbol{\gamma}^{(g)}, \tau^2\mathbf{I})$ and set $\boldsymbol{\gamma}^{(g+1)}$ equal to $\boldsymbol{\gamma}_{can}$ with probability:

$$\alpha(\boldsymbol{\gamma}^{(g)}, \boldsymbol{\gamma}_{can}|\mathbf{Y}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{Z}, \boldsymbol{\zeta}) = \min\left\{1, \frac{\prod_{k=1}^{K}\left\{\left[\frac{\exp(\mathbf{w}_k'\boldsymbol{\gamma}_{can})}{1+\exp(\mathbf{w}_k'\boldsymbol{\gamma}_{can})}\right]^{\zeta_k}\left[1-\frac{\exp(\mathbf{w}_k'\boldsymbol{\gamma}_{can})}{1+\exp(\mathbf{w}_k'\boldsymbol{\gamma}_{can})}\right]^{1-\zeta_k}\right\}p(\boldsymbol{\gamma}_{can})}{\prod_{k=1}^{K}\left\{\left[\frac{\exp(\mathbf{w}_k'\boldsymbol{\gamma}^{(g)})}{1+\exp(\mathbf{w}_k'\boldsymbol{\gamma}^{(g)})}\right]^{\zeta_k}\left[1-\frac{\exp(\mathbf{w}_k'\boldsymbol{\gamma}^{(g)})}{1+\exp(\mathbf{w}_k'\boldsymbol{\gamma}^{(g)})}\right]^{1-\zeta_k}\right\}p(\boldsymbol{\gamma}^{(g)})}\right\}$$

and set $\boldsymbol{\gamma}^{(g+1)}$ equal to $\boldsymbol{\gamma}^{(g)}$ otherwise, where $t_\nu(\boldsymbol{\mu}, \mathbf{T})$ is the multivariate t distribution with mean $\boldsymbol{\mu}$, scale matrix $\mathbf{T}$, and ν degrees of freedom; and $\phi(\cdot)$ is the standard normal pdf. We choose ν and τ^2 to achieve an acceptance rate of approximately 25%. Note that conditional on the latent mixture indicators (the ζ_k's) this update is equivalent to a random walk Metropolis update for a logistic regression model where $\boldsymbol{\zeta}$ is treated as the dependent variable.

References

Albert, James H. 1992. "Bayesian Estimation of Normal Ogive Item Response Curves Using Gibbs Sampling." *Journal of Educational Statistics* 17(Fall):251–269.

Albert, James H., and Siddhartha Chib. 1993. "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association* 88(June):669–679.

Baum, Lawrence. 1988. "Measure Policy Change in the United States Supreme Court." *American Political Science Review* 82(September):905–912.

Berry, S.M., C.S. Reese, and P.M. Larkey. 1999. "Bridging Different Eras in Sports." *Journal of the American Statistical Association* 94:668–679.

Bradlow, Eric T. 2000. "A Bayesian Random Effects Model for Testlets." *Psychometrika* 64(2):153–168.

Chib, Siddhartha. 1995. "Marginal Likelihood From the Gibbs Output." *Journal of the American Statistical Association* 90(December):1313–1321.

Chib, Siddhartha, and Ivan Jeliazkov. 2001. "Marginal Likelihood from the Metropolis-Hastings Output." *Journal of the American Statistical Association* 96(453):270–281.

Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2000. "The Statistical Analysis of Legislative Behavior: A Unified Approach." Paper presented at the Annual Meeting of the Political Methodology Society.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper & Row.

Enelow, James, and Melvin Hinich. 1984. *The Spatial Theory of Voting: An Introduction*. Cambridge: Cambridge University Press.

Epstein, Lee, Valerie Hoekstra, Jeffrey A. Segal, and Harold J. Spaeth. 1998. "Do Political Preferences Change: A Longitudinal Study of U.S. Supreme Court Justices." *Journal of Politics* 60(Augut):801–818.

Epstein, Lee, and Jack Knight. 1998. *The Choices Justices Make*. Washington, DC: CQ Press.

Epstein, Lee, and Carol Mershon. 1996. "Measuring Political Preferences." *American Journal of Political Science* 40(February):261–294.

Epstein, Lee, Thomas G. Walker, and William J. Dixon. 1989. "The Supreme Court and Criminal Justice Disputes: A Neo-Institutional Perspective." *American Journal of Political Science* 33(November):825–841.

Eskridge, William N. 1991. "Overriding Supreme Court Statutory Interpretation Decisions." *Yale Law Journal* 101(November):331–417.

Hartigan, J.A. 1968. "Inference from a Knockout Tournament." *Annals of Mathematical Statistics* 39(April):583–592.

Johnson, Valen E., and James H. Albert. 1999. *Ordinal Data Modeling*. New York: Springer.

Kass, Robert E., and Adrian E. Raftery. 1995. "Bayes Factors." *Journal of the American Statistical Association* 90(June):773–795.

Londregan, John. 2000a. "Estimating Legislators' Preferred Points." *Political Analysis* 8(Winter):35–56.

Londregan, John. 2000b. *Legislative Institutions and Ideology in Chile's Democratic Transition*. Cambridge: Cambridge University Press.

McKelvey, Richard D. 1979. "General Conditions for Global Intransitivities in Formal Voting Models." *Econometrica* September:1085–1112.

McKelvey, Richard D., and Thomas R. Palfrey. 1995. "Quantal Response Equlibria for Normal Form Games." *Games and Economic Behavior* 10(July):6–38.

McKelvey, Richard D., and Thomas R. Palfrey. 1996. "A Statistical Theory of Equilibrium in Games." *The Japanese Economic Review* 47(2):186–209.

McKelvey, Richard D., and Thomas R. Palfrey. 1998. "Quantal Response Equlibria for Extensive Form Games." *Experimental Economics* 1(1):9–41.

Oh, Man-Suk, and Adrian Raftery. 2000. "Bayesian Multidimensional Scaling and Choice of Dimension." Center for Statistics and the Social Sciences Working Paper 10, University of Washington.

Patz, Richard J., and Brian W. Junker. 1999. "A Straightforward Approach to Markov Chain Monte Carlo Methods for Item Response Models." *Journal of Educational and Behavioral Statistics* 24(Summer):146–178.

Poole, Keith T., and Howard Rosenthal. 1991a. "Patterns of Congressional Voting." *American Journal of Political Science* 35(February):228–278.

Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll-Call Voting*. Oxford: Oxford University Press.

Pritchett, C. Herman. 1948. *The Roosevelt Court: A Study in Judicial Political Values, 1937-1947*. New York: Macmillan.

Rohde, David W., and Harold J. Spaeth. 1976. *Supreme Court Decision Making*. San Francisco: W. H. Freeman.

Schubert, Glendon. 1959. *Quantitative Analysis of Judicial Behavior*. Glencoe, IL: Free Press.

Schubert, Glendon. 1965. *The Judicial Mind: The Attitudes and Ideologies of Supreme Court Justices, 1946-1963*. Evanston: Northwestern University Press.

Schubert, Glendon. 1974. *The Judicial Mind Revisited: Psychometric Analysis of Supreme Court Ideology*. London: Oxford University Press.

Segal, Jeffrey A., and Albert D. Cover. 1989. "Ideological Values and the Votes of U.S. Supreme Court Justices." *American Political Science Review* 83(June):557–565.

Segal, Jeffrey A., Lee Epstein, Charles M. Cameron, and Harold J. Spaeth. 1995. "Ideological Values and the Votes of U.S. Supreme Court Justices Revisited." *Journal of Politics* 57(August):812–823.

Segal, Jeffrey A., and Harold J. Spaeth. 1989. "Decisional Trends on the Warren and Burger Courts: Results from the Supreme Court Data Base Project." *Judicature* 73(June-July):103–107.

Segal, Jeffrey A., and Harold J. Spaeth. 1993. *The Supreme Court and the Attitudinal Model*. Cambridge: Cambridge University Press.

Shepsle, Kenneth A. 1979. "Institutional Arrangements and Equilibrium in Multidimensional Voting Models." *American Journal of Political Science* 23(February):27–59.

Signorino, Curtis S. 1999. "Strategic Interaction and the Statistical Analysis of International Conflict." *American Political Science Review* 93(2):279–297.

Spaeth, Harold J. 1990. "The Dimensions of Civil Liberties Decision Making on the Burger Court." In *Studies in U. S. Supreme Court Behavior* (Harold J. Spaeth, and Saul Brenner, editors), New York: Garland.

Spaeth, Harold J. 1999. *United States Supreme Court Judicial Database, 1953-1998 Terms [Computer File]*. Ann Arbor, MI: Inter-University Consortium for Political and Social Research, 15th edition.

Tate, C. Neal, and Roger Handberg. 1991. "Time Binding and Theory Building in Personal Attribute Models of Supreme Court Voting Behavior, 1916-88." *American Journal of Political Science* 35(May):460–480.