

Scaling Politically Meaningful Dimensions Using Texts and Votes

Benjamin E. Lauderdale London School of Economics
Tom S. Clark Emory University

Item response theory models for roll-call voting data provide political scientists with parsimonious descriptions of political actors' relative preferences. However, models using only voting data tend to obscure variation in preferences across different issues due to identification and labeling problems that arise in multidimensional scaling models. We propose a new approach to using sources of metadata about votes to estimate the degree to which those votes are about common issues. We demonstrate our approach with votes and opinion texts from the U.S. Supreme Court, using latent Dirichlet allocation to discover the extent to which different issues were at stake in different cases and estimating justice preferences within each of those issues. This approach can be applied using a variety of unsupervised and supervised topic models for text, community detection models for networks, or any other tool capable of generating discrete or mixture categorization of subject matter from relevant vote-specific metadata.

A fundamental component of contemporary political science is concerned with measuring political actors' preferences. A primary question in this research endeavor concerns the dimensionality of preferences. Qualitatively oriented scholars of legislatures and courts often argue preferences are in fact highly multidimensional, while quantitative analysts generally argue preferences can be described accurately and succinctly as low-dimensional. These arguments are usually made by reference to statistical, rather than substantive, criteria for dimensionality. How many dimensions do we need to fit the data well? Or at least well enough? The answer is usually one dimension, and only rarely more than two dimensions (Clinton, Jackman, and Rivers 2004; Grofman and Brazill 2002; Hix, Noury, and Roland 2006; Poole and Rosenthal 1997; Voeten 2000).

At the same time, qualitative studies of politics often suggest that preferences are highly multidimensional in consequential ways. For example, that Supreme Court justices' preferences vary across substantive areas of the law has affected the extent to which individual members of the Court are pivotal (e.g., Lauderdale, and Clark 2012) and can influence the direction of the law (e.g.,

Jeffries 1994; Stern and Wermiel 2010). In other contexts, multidimensional preferences give rise to well-known social choice pathologies, and the study of political institutions has been concerned with how those institutions help resolve those problems. Unfortunately, because scholars increasingly focus on low-dimensional preferences, we often miss these important theoretical puzzles. What is more, a focus on low-dimensional preferences limits our ability to answer central puzzles in politics. For example, does a change in the effective dimensionality of a voting body induce change to supposedly stabilizing institutions? These and other limitations suggest a need for the ability to study high-dimensional preferences in political institutions.

Two major obstacles have prevented scholars from moving beyond two dimensions. First, the tools scholars use to reduce voting behavior into latent ideal point estimates do not provide any information about the substantive interpretation of recovered dimensions. As a consequence, the analyst must specify how the estimated dimensions are related to interpretable political cleavages. This is difficult to do because it requires implicitly or explicitly constraining all possible linear transformations of

Benjamin E. Lauderdale is Associate Professor, Department of Methodology, London School of Economics, Houghton Street, London, England WC2A 2AE (b.e.lauderdale@lse.ac.uk). Tom S. Clark is Asa Griggs Candler Professor, Department of Political Science, 327 Tarbuton Hall, Emory University, Atlanta, GA 30322 (tom.clark@emory.edu). We thank Drew Linzer, Phil Schrodt, Jeff Staton, and Kevin Quinn for helpful comments and discussions. We thank the National Science Foundation for financial support (SES-0909235). All replication materials have been deposited in an AJPSS Dataverse and can be located at <http://dx.doi.org/10.7910/DVN/22816>.

American Journal of Political Science, Vol. 00, No. 0, xxx 2014, Pp. 1–18

© 2014 Midwest Political Science Association

DOI: 10.1111/ajps.12085

the latent space, and the number of possible linear transformations grows rapidly with the dimensionality (Rivers 2003). Second, the binary data most commonly used to estimate preferences (votes) are insufficiently informative to support analyses beyond one or two dimensions using multidimensional scaling methods. Improvements in fit tend to fall off very quickly, not necessarily because the political world has only one or two dimensions, but because the limited voting data we have can be predicted well enough by one or two estimated dimensions, given the number of voters and votes that are typically present. Political scientists using ideal point estimates have therefore focused heavily on these highly simplified measures of preferences: just one or two measured dimensions that may combine a wide range of substantive political issues across which preferences are correlated, but not perfectly so. The fundamental problem is that it is not usually possible to figure out which of these underlying substantive issues a given vote is about, if one looks at nothing more than the observed pattern of votes.

We introduce a generally applicable method for resolving these two problems. The key innovation is that we bring metadata about votes to bear on the problem of describing legislative and judicial behavior. While different kinds of metadata can be used for this purpose, in this article, we focus on the use of vote-specific texts (e.g., bills, opinions, debates) to identify which of many issue dimensions a given vote was about. This is in contrast to standard models that try to infer this information from the pattern of roll-call voting behavior. By using metadata about which votes are similar in substance, we can statistically identify multidimensional estimates that improve fit, even when appropriately penalized for the additional uncertainty that results from estimating additional parameters.

A range of models exist for modeling texts as mixtures of, or as discrete members of, topics, under the assumption that different topics have different characteristic distributions of term usage. As we show in the next section, these discrete or mixture categorizations can be used to solve an identification problem in multidimensional ideal point estimation. Ideal point estimators effectively discover which legislators or judges tend to vote together, but not what they are voting about. In contrast, topic models discover the topic or topics being discussed in a series of texts, but not what is being said about those topics. Roughly speaking, combining the two allows us to discover what issue a vote is about from textual data and how legislators or judges respond to that issue through their votes. In this article, we demonstrate this combination using latent Dirichlet allocation (LDA) for textual data (Blei, Ng, and Jordan 2003) and item response the-

ory (IRT) models for ideal point estimation from roll-call data (Jackman 2001). In the third section, we describe the technical details of how we combine Markov chain Monte Carlo (MCMC) simulation of the LDA and IRT models, targeting optimal prediction of the roll-call votes rather than the texts. We recover a list of issues/dimensions, each with characteristic terms and spatial preference estimates.

In the fourth section, we apply our model to the U.S. Supreme Court, using the texts of opinions as our source of information about the issues at stake in each decision. In the data we consider, we find the best fit under models with 20–30 issue dimensions. These issue dimensions are substantively meaningful (i.e., search and seizure, patent and copyright) and are discovered from the textual data. Identifying the dimensions this way greatly reduces the issues of identification and interpretation generally associated with standard multidimensional ideal point models. In the conclusion, we discuss a range of alternative methods that could be used to fulfill the role that the LDA model plays in this article. The same approach can be applied using any method of calculating topic assignments: supervised or unsupervised, with discrete or mixture assignments, using textual or other data that convey information about topical similarity.

Combining Texts and Votes

Scholars have increasingly argued that additional information is required to make sense of spatial preference estimates in a variety of contexts, ranging across the European Parliament (Hix and Crombez 2005), the U.S. Congress (Crespin and Rohde 2010), the Brazilian legislature (Zucco and Lauderdale 2011), and the U.S. Supreme Court (Lauderdale and Clark 2012). The goal of leveraging metadata is to overcome the limitations of multidimensional scaling models, which only recover dimensions in voting behavior that maximize statistical fit, even when those dimensions conflate many substantive dimensions scholars want to study. Metadata about the subject of particular votes have the potential to provide the estimator with information about the substantive meaning of the dimensions.

To understand the intuition, consider a comparison between scaling models in political science and in educational testing. Major educational tests like the SAT and ACT are split into sections—math, verbal, and so forth—based on the test designers’ knowledge of what the questions are about. This is viewed as preferable to attempting to recover multiple traits from an undifferentiated

collection of test items because it does not require the scaling model to be able to differentiate the questions into different topics. By contrast, traditional multidimensional scaling in political science is akin to educational tests in which the verbal, math, and other questions are collectively scaled in two (or more) dimensions, without any labeling of which questions are expected to test which abilities. The consequence is uncertainty about what is being measured, which is avoided in the educational testing context by writing separate tests that each aim to test primarily a single ability of interest. Fortunately, while political actors do not usually provide researchers tidy, discrete categorizations of which votes are about which issues, they do provide metadata information—texts of bills, opinions, oral arguments, and legislative debates; citation data to legal statutes and precedent cases; and so forth—that can help us figure out what their votes were about and what preference dimensions those votes reveal. While manual coding of votes into discrete categories is a viable approach in some applications, political scientists often examine large sets of votes, making automatic processing of metadata valuable.

We focus on textual metadata in this article. The most widely analyzed quantitative feature of texts—relative usage of different terms—tends to reveal a great deal about what issues are being discussed. For example, consider the Supreme Court case *Florence v. Board of Chosen Freeholders of the County of Burlington*, decided in 2012. This case considered whether police need to have specific reasons to suspect contraband in order to conduct a strip search of an individual who is already being jailed. The majority opinion repeatedly uses terms like *jail* (50 appearances), *search* (45), *detainee* (37), *inmate* (28), *contraband* (27) and *drug* (14), which also appear frequently in some other Court decisions but appear rarely in most opinions. One can infer that the subject of the *Florence* case is more similar to opinions that use terms like *detainee* and *search* than opinions that do not use those terms. Of course, there are other terms like *petitioner* (23) that are much more generic features of Supreme Court opinion texts and provide little information about which issues are at stake. Using a statistical model of text, we can determine a mix of issues implicated in the opinions. Our approach relies on the assumption that this mix also reveals which *preference dimensions* are relevant to this vote. A text that is purely about a single issue is assumed to imply a vote that only depends on preferences on that issue.

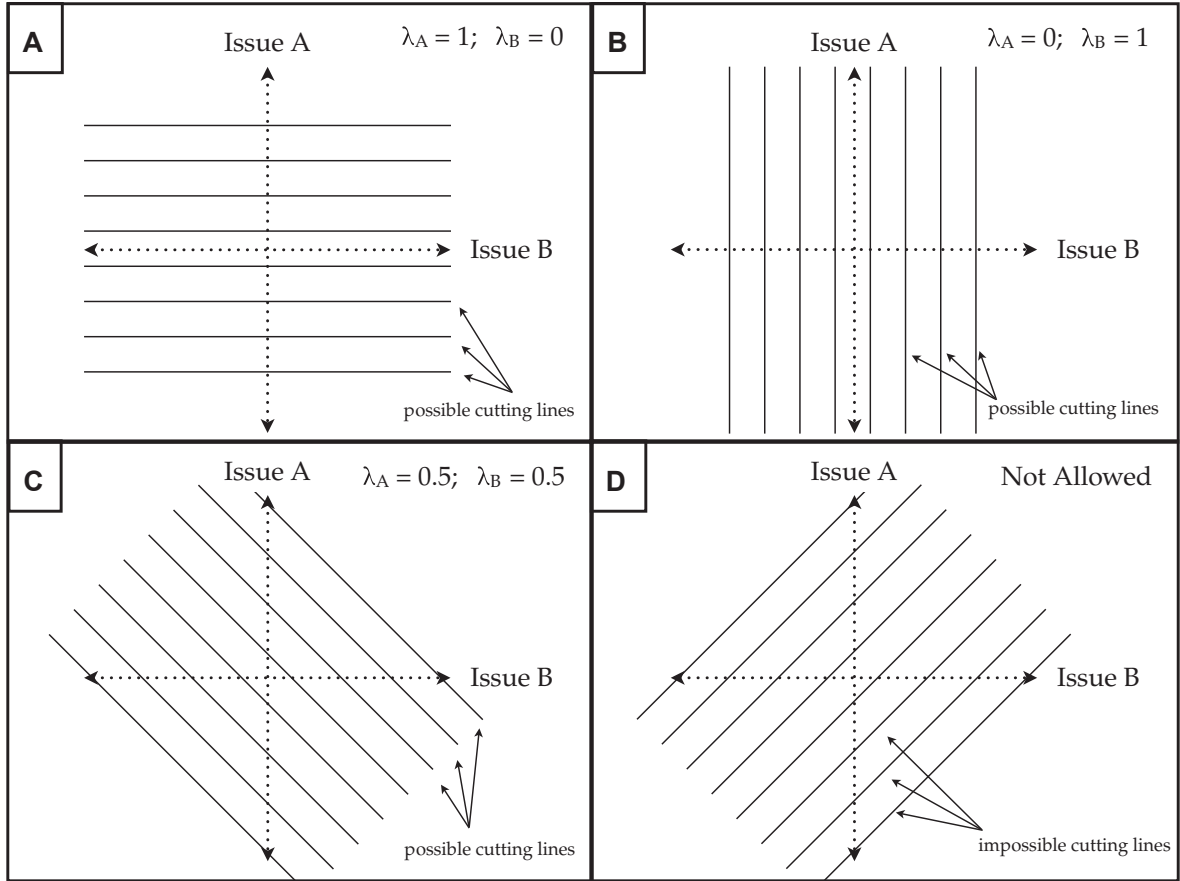
Consider a simple example, with just two issue dimensions. Voter i has a position θ_{iA} on issue A and θ_{iB} on issue B. The question for any given vote is this: to what extent is that vote about issue A and to what extent is

that vote about issue B? Our approach is to use the text to estimate the answer. Specifically, let λ_A represent the degree to which the vote is about issue A and λ_B represent the degree to which the vote is about issue B, where $0 \leq \lambda_A, \lambda_B \leq 1$, and $\lambda_A + \lambda_B = 1$. Figure 1 shows several special cases for this two-dimensional spatial model. Panel A shows possible cutting lines—the lines separating yes voters from no voters—if the vote is entirely about issue A ($\lambda_A = 1, \lambda_B = 0$). Different cutting lines will correspond to different coalitions of yes and no voters, but only θ_{iA} will predict the behavior of any voter i . On such votes, preferences with respect to issue B are irrelevant to voting behavior. Panel B shows the corresponding case for votes that are purely on issue B.

Panel C shows the case of a bill that is equally about issue A and issue B ($\lambda_A = 0.5, \lambda_B = 0.5$). We assume that the weight each dimension plays in each vote is *directly proportional* to the proportion of text coming from each issue. We measure a vector λ_j such that each element $0 \leq \lambda_{jd} \leq 1$ and $\sum_{d=1}^D \lambda_{jd} = 1$ from the text, and then apply these relative weights to voter i 's position θ_{id} across dimension d in determining her latent voting preferences on vote j . These are nontrivial identifying assumptions: they determine how observed votes that mix multiple issues shape our estimates of preferences across those issues.

Panel D shows cutting lines that are impossible under the model. These are cutting lines that would describe coalitions of issue A conservatives with issue B liberals opposing coalitions of issue A liberals with issue B conservatives. While such coalitions are, of course, conceptually and politically possible, ruling them out by assumption is a consequence of using a vector of mixture weights λ to identify the rotation of the cutting line. A valuable side effect is that this restriction on which cutting lines—more generally, hyperplanes—are allowed enforces consistent liberal/conservative labeling across dimensions. Nearly all of the possible rotations and reflections of the space that do not change predicted voting patterns under a standard multidimensional IRT model (Rivers 2003) *do* change predicted voting patterns under our model because of this restriction.

In principle, we could estimate the orientation of the cutting hyperplanes for each vote using any of a variety of models to map textual data about votes into vectors of mixture components λ_j . LDA is well suited to generating suitable λ_j , using minimally processed text and without prior specification of topics. Many other approaches are possible, including discrete topic models for text or expert issue codings, that would only allow a single preference dimension to predict each vote.

FIGURE 1 Cutting Lines for Several Mixtures of Two Topics

Note: The top left panel (A) shows possible cutting lines that can be estimated given a text that is entirely drawn from issue A. The top right panel (B) shows possible cutting lines that can be estimated given a text that is entirely drawn from issue B. The bottom left panel (C) shows possible cutting lines that can be estimated given a text that is equally drawn from issues A and B. The bottom right panel (D) shows cutting lines that are impossible under the model.

Methods

Model Specification

Our model has two major components. We first estimate the latent distribution of topics across and within a series of documents using LDA. We then use the posterior estimates of the distribution of topics for each text from the LDA model to model the extent to which the vote on that text was about each dimension. The LDA portion of the model is identical in specification and implementation to previous MCMC implementations (Griffiths, and Steyvers 2004; Heinrich 2009); however, we change some notation for compatibility with the notation of the voting model. The notation for indices, parameters, and data in the model is listed in Table 1.

The LDA model terminology refers to *words*, *terms*, *topics*, and *documents*. A document is considered a sequence of words. Each word is an instance of a particular

term. For example, the sentence “The Republican and the Democrat disagreed” has six words and five terms (the term *the* appears as both the first and fourth word). LDA assumes the relative frequency of different terms in different documents reveals which topics are being discussed. The model can be summarized as follows. Each word that appears in a document is assumed to have a latent topic z_{jk}^* for word k in document j is assumed to be a multinomial draw from the distribution of topics in that document: $z_{jk}^* \sim \mathcal{M}(\lambda_j)$. The observed term w_{jk} for word k in document j is assumed to be a multinomial draw from the distribution of terms in the latent topic z_{jk}^* : $w_{jk} \sim \mathcal{M}(\varphi_{z_{jk}^*})$. The multinomial parameter vector describing the distribution of terms t within each topic d is assumed to be a draw from a Dirichlet distribution with hyper parameters δ : $\varphi_d \sim \mathcal{D}(\delta)$. The multinomial parameter vector describing the distribution of topics d within each document j is

TABLE 1 Model Notation

Indices	$i \in 1, 2, \dots, N$	Index for voters
	$j \in 1, 2, \dots, M$	Index for votes/documents
	$d \in 1, 2, \dots, D$	Index for dimensions/topics/issues
	$k \in 1, 2, \dots, L_j$	Index for words in text for vote j
	$t \in 1, 2, \dots, V$	Index for terms in vocabulary
Latent and observed data	$w_{jk} \in 1, 2, \dots, V$	Observed term for each word k in document j
	$z_{jk}^* \in 1, 2, \dots, D$	Latent topic of each word k in document j
	$y_{ij} \in 0, 1$	Vote j of voter i
	y_{ij}^*	Latent vote j of voter i
	Φ_{td}	Mixture component of term t in topic d
Parameters to be estimated	λ_{jd}	Mixture component of topic d in document j
	α_j, β_j	Discrimination parameters for each vote j
	θ_{id}	Ideal point for voter i on topic d
	ρ	Prior correlation of ideal points across topics

assumed to be a draw from a Dirichlet distribution with hyperparameters $\boldsymbol{\gamma}$: $\boldsymbol{\lambda}_j \sim \mathcal{D}(\boldsymbol{\gamma})$. The model assumes that a set of topics describes variation in term use across the corpus of documents: the topic for each word location in each document is a multinomial draw from a distribution of topics characteristic of that document, and the actual term that appears in that word location is a multinomial draw from the distribution of terms characteristic of that topic.

We are particularly interested in the $\boldsymbol{\lambda}_j$, the vectors of mixture components describing the fraction of the text for each vote $j \in 1, 2, \dots, M$ that is in each topic $d \in 1, 2, \dots, D$. We are also interested in the term frequency distributions $\boldsymbol{\varphi}_d$ for each topic, since these allow us to substantively label the topics by calculating which terms are most indicative of a word being drawn from each topic.

In D dimensions, the Jackman (2001) model for observed votes y_{ij} has the following form:

$$y_{ij} = \begin{cases} 1 & \text{if } y_{ij}^* \geq 0 \\ 0 & \text{if } y_{ij}^* < 0, \end{cases} \quad (1)$$

$$y_{ij}^* \sim \mathcal{N}\left(\alpha_j + \sum_{d=1}^D \beta_j \lambda_{jd} \theta_{id}, 1\right). \quad (2)$$

Our model is a restriction of this model: we reduce β from a vector of length D for each vote j to a scalar for each vote j . In the conventional multidimensional IRT model, the vector of discrimination parameters describes the degree to which each latent dimension predicts the latent outcome. In our model, the *relative* degree that each latent dimension predicts the latent outcome is determined

by the text, and only the degree to which the resulting weighted average of the dimensions predicts the outcome is estimated from the voting data. From the LDA model on the texts, we get a D -dimensional vector $\boldsymbol{\lambda}_j$ describing the fraction of vote j 's text that comes from each topic d . This vector determines the relative weight of each ideal point dimension in determining the latent vote:

$$y_{ij}^* \sim \mathcal{N}\left(\alpha_j + \sum_{d=1}^D \beta_j \lambda_{jd} \theta_{id}, 1\right). \quad (3)$$

This can be rewritten in terms of vote-specific one-dimensional ideal points θ_{ij} :

$$y_{ij}^* \sim \mathcal{N}(\alpha_j + \beta_j \theta_{ij}), \quad (4)$$

$$\theta_{ij} \equiv \sum_{d=1}^D \lambda_{jd} \theta_{id}. \quad (5)$$

The θ_{id} are issue-specific ideal points for each voter, which each correspond to preferences on a vote with corresponding text entirely on a single topic/issue. Combined with the estimated mixture of topics $\boldsymbol{\lambda}_j$ for the document associated with a particular vote j , one can use these issue-specific ideal points to calculate the vote-specific ideal point θ_{ij} .

Model Priors and Simulation

For the LDA model, the choice of priors determines the symmetry and sparsity of the recovered topics. We use symmetric Dirichlet priors, which yields topics of roughly equal total text length. While this may not be ideal for all

topic-modeling applications (Wallach, Mimno, and McCallum 2009), it ensures that roughly equal quantities of voting data are brought to bear on estimating each dimension. Having chosen symmetric priors, we still must choose the values of the hyperparameters γ and δ . Smaller values of γ lead to more decisive assignments of documents to a small set of topics. Smaller values of δ lead to more decisive assignments of terms to a small set of topics. We are more interested in decisive assignment of documents to topics than terms to topics, so we assume different values than those recommended in the literature (Griffiths, and Steyvers 2004). We set the hyperparameter for the assignment of topics within documents equal to $\gamma = 1/D$ and the hyperparameter for the assignment of terms to topics $\delta = 1$.

For the IRT model, we use normal priors with mean 0 and variance 4 for the vote parameters α_j and β_j : $\alpha_j \sim \mathcal{N}(0, 4)$ and $\beta_j \sim \mathcal{N}(0, 4)$. For the ideal points, we assume a multivariate normal prior with mean 0, and a covariance matrix with 1 on the diagonal and $\rho \in (0, 1)$ off the diagonal. The mean of 0 centers the ideal point scale, and the diagonal components of the variance matrix define the spread of the ideal point scale. The positive off-diagonal elements of the variance matrix orient all the issue-specific dimensions with conservatives at the same ends and provides shrinkage of ideal points in each dimension to their common mean across dimensions. We put a uniform prior over positive values of this correlation: $\rho \sim \mathcal{U}(0, 1)$.¹

We estimate our model by MCMC simulation. Our estimator is built from the LDA collapsed Gibbs sampler described by Griffiths and Steyvers (2004) and the IRT Gibbs sampler described by Jackman (2001). Rather than estimating a fully joint model we simulate the LDA model, and then condition the IRT model on the current draw of $\boldsymbol{\lambda}$ from the LDA model. We describe our sampling procedure in the online appendix.

Model Selection

Choosing the best number of topics for a topic model is a difficult problem in general, and several additional considerations are introduced by our application of the topic mixture vector $\boldsymbol{\lambda}$ to predicting votes. A common approach to assessing the best number of topics in LDA modeling is to assess the model's *perplexity* under a number of topics (Blei, Ng, and Jordan 2003). Perplexity is a measure that evaluates how well the model predicts held-out textual

data. This criterion is used to find when adding more topics stops making appreciable increases in the model's predictive power; however, such a criterion is only appropriate if the goal is predicting the textual data. Our goal, by contrast, is predictions of roll-call votes. If we estimate too few topics, we will fail to find variation in preferences across issues that could improve the predictive power of the ideal point estimates. If we estimate too many topics, we will have only a few votes on each issue and will lack sufficient data to learn much about preferences on that issue. To balance these concerns about having enough votes per topic with having enough topics to adequately describe the voting space, we use the deviance information criterion (DIC; Spiegelhalter et al. 2002) to optimize predictive power of the ideal points for votes. Because we estimate the LDA and IRT models in two steps, we can easily calculate the DIC on the basis of only the latter. Where the deviance $D(\Theta) = -2 \log \mathcal{L}(\Theta)$, we calculate the DIC as two times the mean posterior deviance minus the deviance of the mean posterior estimates of the parameters:

$$DIC = 2\overline{D(\Theta)} - D(\bar{\Theta}). \quad (6)$$

Model Features and Limitations

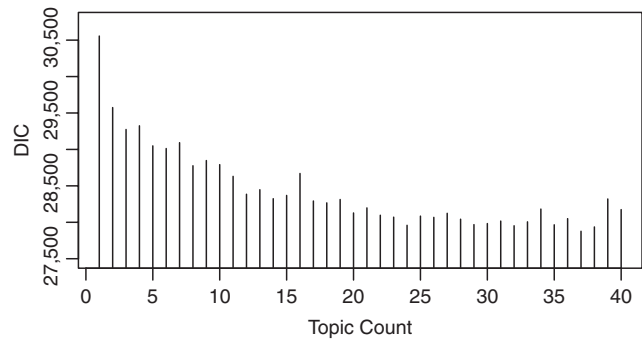
Our model assumes that all voters are voting on the same mixture of dimensions: if 35% of the text is from topic A and 65% from topic B, then every voter's preference is a 35–65 weighted average of his or her preferences on dimensions A and B. This assumption is required for model identification, but there are two possible substantive objections in our application. First, the proportions of text are surely not identical to the relative importance that the justices put on each of the two dimensions in making their decisions. Second, the justices may differ in which dimension they think is most relevant to a given case. Unfortunately, if we explicitly allow justices to respond to different mixtures of dimensions, by using only the opinion texts that a given justice wrote or joined, the preference scales we estimate will no longer be comparable across justices. However, even if justices really are responding to somewhat different dimensions, our approach is a reasonable approximation. This is also an argument in favor of assigning cases to mixtures of topics, rather than using a discrete topic model or a discrete expert coding: if the justices talk about multiple issue areas in their opinions, allowing their preferences in all those areas to predict their decisions is a more reasonable compromise than identifying just one area.

¹ Note that a more flexible covariance structure between dimensions could be estimated if a large number of voters were available.

There are two senses in which the estimates we derive under this approach are not unique. First, as noted previously, the LDA model combined with text is only one way to generate a matrix of weights λ_{jd} for each vote j across each dimension d . Each method for defining λ_{jd} will define a different multidimensional preference space. This is sufficient to meet the identification restrictions identified by Rivers (2003); however, it is vital that one uses information that is appropriate for defining a meaningful, multidimensional political space. This is more or less guaranteed if one uses a type of metadata that immediately provides a valid $\mathbf{\lambda}$ matrix. For the Supreme Court, the Spaeth issue and issue area codes provide an expert, discrete issue coding, which we discuss in comparison to our estimated topics later in this article. For legislatures, originating committee is often used as a topic coding. For textual and other sources of metadata, it is important to have a substantive justification for why the chosen metadata model provides a useful and meaningful mapping of votes onto dimensions. In our application below, we provide this justification with reference to the limitations of the Spaeth codes by examining the details of the recovered topics and by demonstrating that the estimates generate novel insights into preference variation on the Court.

Second, even once one has decided to fit a topic model on a particular set of texts to define the λ_{jd} , it is important to recognize that such models do not yield “definitive” topical decompositions (Blei and Lafferty 2009). The idea of “topics” is useful for characterizing variation, but there are no “true topics” to discover in the data. Thus, there is no generally “correct” answer to the dimensionality issue. The performance of the models in terms of fit (as assessed by the DIC) will vary to the extent that the topical decomposition tracks the underlying variation in the preferences of the voters. Moreover, when LDA is estimated by MCMC as in this article, it does not generate exactly the same topics each time it is reestimated. The appropriate way to think about this is as a descriptive model: we find a set of topics that explains variation in term usage, but those topics are not fixed. For some applications, having a fixed-topic coding scheme is vital, and so a supervised topic model for text should be used (Blei and McAuliffe 2007; Hopkins and King 2010). However, for exploratory analyses like the one we describe below, an unsupervised topic model works well and does not demand any *ex ante* specification of political issues. Analysts must exercise substantive judgment about the political spaces they wish to use to describe preferences; they must choose whether they want to generate a space via an exploratory tool like LDA or instead use some other method that allows greater control over the recovered space.

FIGURE 2 DIC for Predicting Roll-Call Votes as a Function of Topic Count



Application: The U.S. Supreme Court

The model we develop above is broadly applicable to many political institutions. The voting data modeled with the IRT model must be linked to documents that vary in term usage primarily as a function of the issues at stake in the vote. Such texts are not hard to find: legislators often have recorded debates about proposed laws, courts write opinions summarizing the issues and legal questions associated with the particular case they are deciding, and bodies like the United Nations vote on resolution and treaty texts. To illustrate the utility of our model, we employ data from the U.S. Supreme Court. Our roll-call data consist of 4,321 nonunanimous Supreme Court decisions from 1949 to 2006. These include the votes of 30 justices, of whom no more than nine were involved in any single decision. The text we associate with each decision is that of the opinions in a case.² We employ the full text of opinions, rather than isolating specific aspects, such as the holding, because we believe the larger content of the opinion is important for establishing the mix of issues considered by the Court.

In order to fit the model, the first question we must consider is the dimensionality of the model we wish to estimate: how many topics for the LDA model best fit the voting data? To roughly estimate the appropriate number of dimensions, we ran short simulations—2,000-iteration burn-in, 4,000-iteration sample—for a range of topic counts from 1 to 40 to find which numbers of topics lead to the best prediction of votes by the DIC as described above. Figure 2 shows the results of these simulations. We find improvement in the DIC up to around 25 topics,

²If one were interested in making predictions about future cases, it would make sense to use legal briefs, appeals court majority opinions, or other texts that are available in advance of the case decision.

and little improvement beyond. The predictive power of the model does not get worse either, primarily due to the cross-dimensional shrinkage of the ideal points.

There is no “correct” number of dimensions. We select 24 dimensions for further analysis because that number seems to provide a reasonable point at which further marginal decreases in the DIC are small. Just as importantly, when we look at the terms that are most characteristic of each topic, this is also around the point at which additional substantive nuance diminishes. Making this evaluation is not amenable to a statistical criterion, but we can see the principle in action by considering the topics recovered when we fit a two-dimensional model. As Figure 2 shows, this model represents a significant improvement in fit over the one-dimensional model. The two topics recovered from the texts describe the two major functions of the Supreme Court: some of the terms that are most indicative of the first dimension are *act*, *congress*, and *federal*, whereas some of the terms that are most indicative of the second dimension are *trial*, *jury*, and *evidence*. The first dimension captures constitutional review of the legislative and executive branches: acts of Congress, issues of federalism, taxes, etc. The second dimension captures oversight and administration of the judiciary: criminal defendants’ rights, rules of criminal and civil procedure, etc. However, as we show below when we examine the 24-dimensional fit, we can recover many more than two substantively meaningful topics, and as the DIC above indicates, this improves predictive performance. In other words, a combination of both statistical fit and substantive inspection should guide selection of dimensions.

Estimated Topics versus Expert Coding

Unlike most potential applications of our model, for the Supreme Court we have an existing expert coding to compare to the LDA model’s assignment of texts to topics. Issue area codings are available from the Supreme Court Database (Spaeth et al. n.d.). Figure 3 is a heat map showing the distribution of Spaeth issue areas across LDA topics. We adopt the common practice of using the top three terms for each topic as a label. For example, *union*, *labor*, and *board* are the most distinctive terms in a topic consisting of text related to union regulation.³ While our 24 topics are necessarily distinct from the 13 issue areas, there is much in common. Relatively unpopulated Spaeth issue areas like Unions, Interstate Relations, Federal Taxa-

³We do not stem the text, so some groups of terms are duplicative. With shorter texts, stemming may be necessary to get sufficient information from the texts.

tion, and Miscellaneous fall heavily into single LDA topic areas, whereas more common issue areas involve cases that mix across multiple topics.

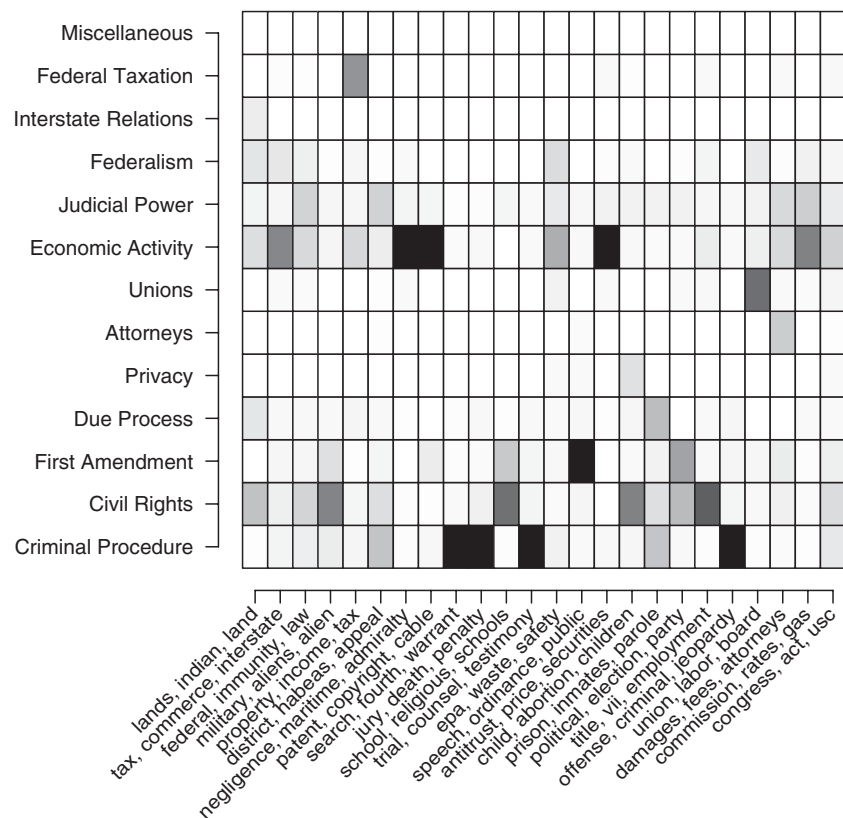
These estimated topics turn out to be superior for predicting variation in justices’ voting behavior. We compared the DIC from estimates based on a 13-dimensional LDA topic model to the DIC from estimates using the Spaeth issue area codings to define the λ matrix and found that the former was lower. We then converted these 13-dimensional topic estimates from the LDA model into a discrete topic coding by assigning each case j to the category d that had the largest estimated λ_{jd} . Using the discretized LDA topic assignments, there was still an improvement in DIC versus the Spaeth assignments, but not as large. Thus, in a comparison of 13 dimensional models, some of the gain in prediction versus Spaeth is due to using texts rather than the Spaeth coding scheme, and some is due to allowing cases to be assigned to mixtures of topics.⁴ Further gains in prediction then arise from moving from 13 to 24 topics.

The fact that some of our gains in predictive power are due to the fact that cases can be assigned to multiple topics, with estimated weights, highlights one of the advantages of our approach versus expert coding. In matters as complex as litigation or legislation—the matters on which we often scale votes—preferences over multiple issues may influence decisions. Depending on coding rules, an expert coder might be forced to choose a single topic to assign a vote, might have the option of selecting multiple issues without specifying their relative weight, or might have the option to assign weights to topics. In the former two scenarios, we do not have the benefit of knowing how relevant different topics are; in the final scenario, the coder has a very difficult task and reliability may be low. The text-based approach avoids these pitfalls by employing a clearly specified model of topic mixtures and relying on straightforwardly measured information: the distribution of terms in documents. This is not to say that the expert-coded topics convey no information, only that the gains they provide over a unidimensional model are not as large as the gains from the mixture topics we recover from the opinion texts. If no texts were available, using this kind of discrete coding to define the λ_{jd} would be a good way to proceed.

The Supreme Court in 24 Dimensions

Figure 4 shows the relative fraction of text in each LDA topic over time. Some topics—“negligence,” “maritime,”

⁴Unfortunately, we cannot do a comparison to a mixture, expert coding, as the Spaeth coding is discrete.

FIGURE 3 Spaeth issue areas versus LDA issues.

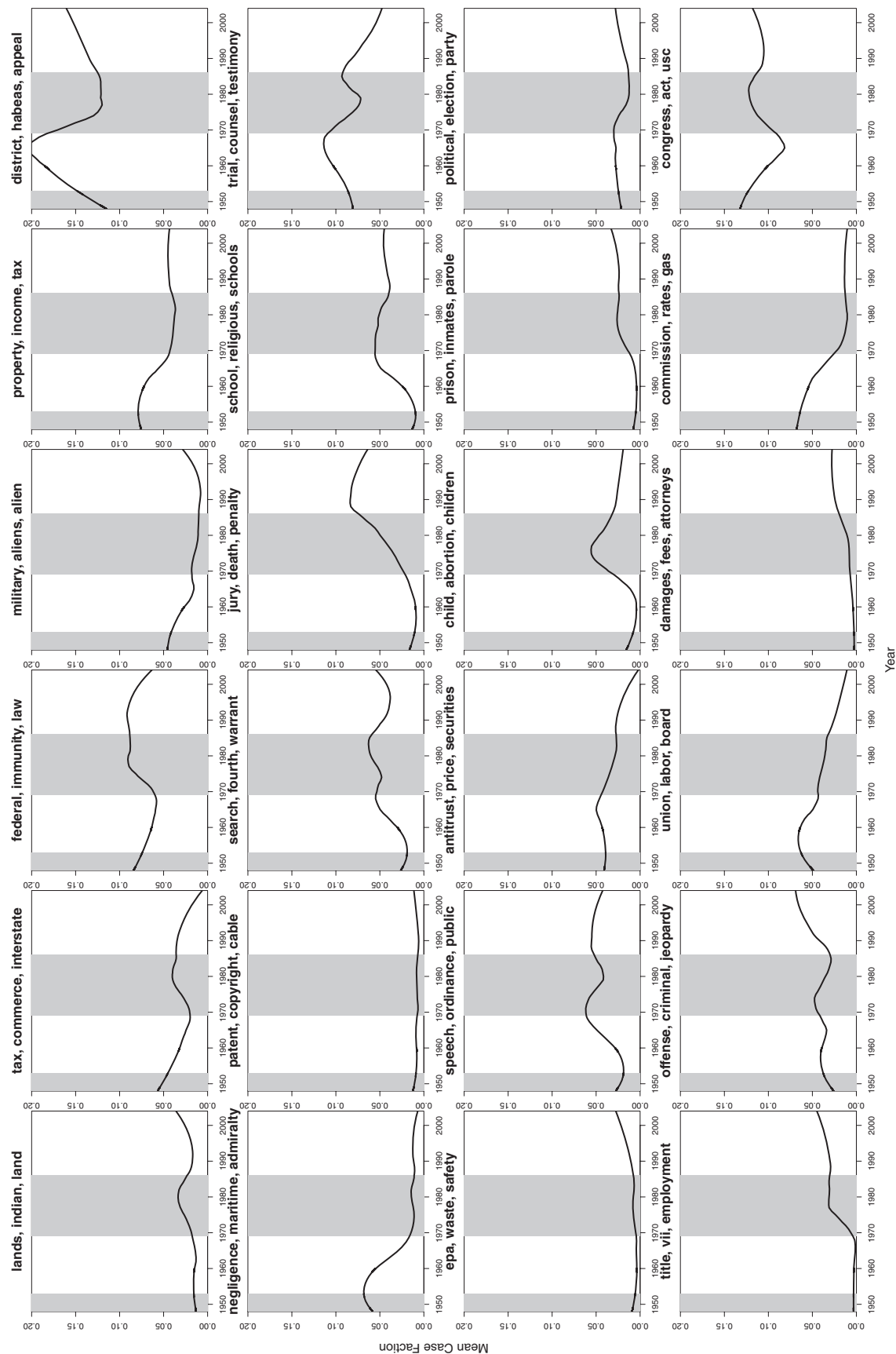
Note: The figure shows the frequency with which each Spaeth issue area is attached to each LDA topic. Darker shades correspond to more frequent appearance.

“admiralty,” and “commission,” “rates,” “gas”—are far more prevalent early in the period, whereas others appear only after specific legislative or legal events—“title,” “vii,” “employment,” and “child,” “abortion,” “children.” Consider cases dealing with the death penalty (the topic associated with the terms *jury, death, and penalty*). We see a relatively low level of representation by this topic during the early period here, followed by a steady increase from about 1970 through 1990, and then a subsequent leveling off or even decrease over the next 15 years. This pattern is consistent with qualitative studies of the death penalty, which document increased litigation during the 1970s and 1980s, followed by a slight decrease during the 1990s (e.g., Walker 2009). As another example, consider the topic related to abortion cases (the topic associated with the terms *child, abortion, and children*). We see a spike in this topic during the 1970s, followed by a steady decrease from 1980 through 2000. This pattern comports with other studies of doctrine that show in the wake of *Roe v. Wade* (1973), there was a lot of Supreme Court litigation that sought to clarify and extend the rule announced in *Roe*, whereas the Supreme

Court has paid less attention to this line of doctrine since the mid-1980s (e.g., Clark and Lauderdale 2012). As a final example, consider the topic associated with the terms *district, habeas, and appeal*. We see here a spike during the Warren Court era (1953–1968), followed by a sharp decrease during the Burger Court (1968–1986), and then a slower but steady increase during the Rehnquist Court years (1986–2005). This pattern comports with conventional understandings—namely, that the Warren Court spent a lot of time dealing with the rights of criminal defendants, that the Burger Court turned its attention away from such cases, and that the Rehnquist Court again returned to these cases (albeit from a different ideological perspective than that of the Warren Court).

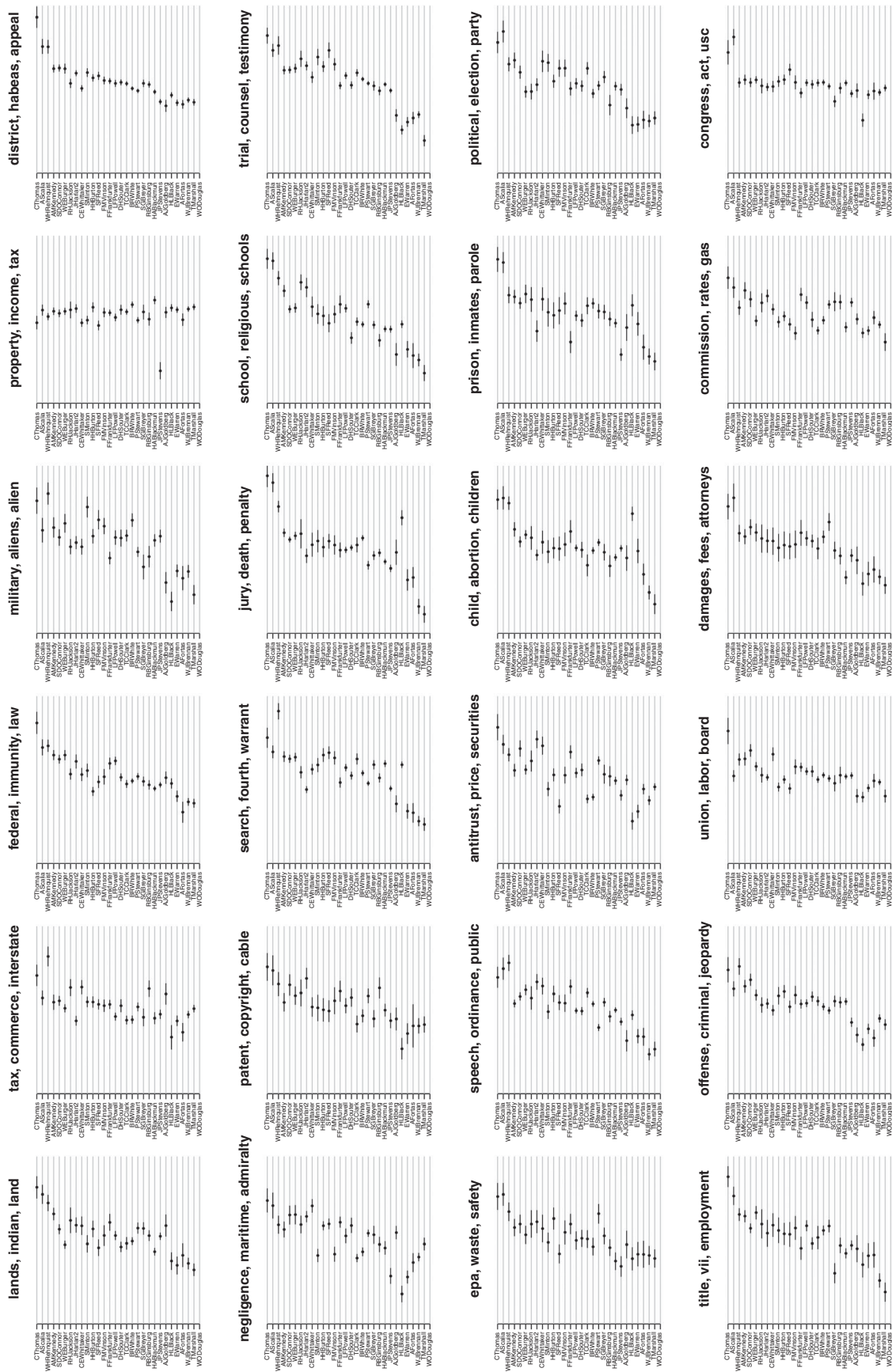
Turning from the content of the Supreme Court’s docket to the preference estimates, in Figure 5, we show the relative positions of the 30 justices in our data, in each issue area. Justice Douglas is not visible on the plots, as he is so far off the left edge of the scale that including him would obscure the more relevant differences between other justices. The justices are listed from top to bottom in decreasing average ideal points across all

FIGURE 4 Frequency of Issue Appearance over Time



Note: Each panel shows the frequency of appearance for each issue over time. The title of each panel is the top three most distinctive terms for the issue. Gray regions identify the chief justice. The four chief justices included in these data are, chronologically, Vinson, Warren, Burger, and Rehnquist.

FIGURE 5 Justice Locations for Each Issue



Note: Each panel shows the posterior mean and standard deviation for each justice, in each issue. The title of each panel is the top three most distinctive terms for the issue. Justices are sorted on their average positions across all dimensions. Justice Douglas's estimated positions are to the negative extreme of each dimension and are excluded to maintain legibility for the remaining justices.

issue dimensions. The justices' rank orderings across issue dimensions are correlated, but not perfectly so. The preferences in two tax issues—"property," "income," "tax" and "tax," "commerce," "interstate"—are very weakly related to preferences in the other dimensions. We also find that the category "congress," "act," "usc", which contains generic terms related to all Court reviews of congressional statutes, has an ordering that is weakly related to other issues. The more that the text of a case's opinions comes from this category, the more that text is about abstract issues related to judicial review in general, and so it is perhaps unsurprising that the typical preference ordering of the Court applies less and less in these cases. In these three areas, it is not just the case that the orderings of justices are atypical, but it is also the case that we cannot be very confident of any ordering. There is not much difference among the justices on these issues. This result is one that is alluded to among scholars of the courts, though rarely documented (but see Martin and Quinn *n.d.*, 4).

The model recovers a number of features that one should expect to find, providing evidence of the model's validity. For example, as noted, Justice Douglas is the leftmost justice on all dimensions, whereas the two most conservative justices on most dimensions are Justices Thomas and Scalia (they are the top two justices depicted in each panel of Figure 5). One exception is the topic associated with the terms "property," "income," and "tax," which does not exhibit much variation in the justices' expressed preferences at all. However, the model does more than just recover well-known ideological consistencies among extremist justices; there are many examples of individual justices with peculiar views in particular issue areas. For example, it has been previously documented that Justice Black had more conservative views on issues of criminal rights than on other issues (e.g., Lauderdale and Clark 2012; Newman 1997), and this pattern is recovered here, as is apparent in the topic "search," "fourth," "warrant;" the topic "prison," "inmates," and "parole" and the topic "jury," "death," "penalty." Taken together, the patterns depicted in Figures 3–5 show that our model is capturing substantively meaningful dimensions of judicial decision making and substantively relevant variation in the justices' preferences across those dimensions.⁵

Figure 6 shows the location of the median for each topic for each of the 57 years covered by our data. These

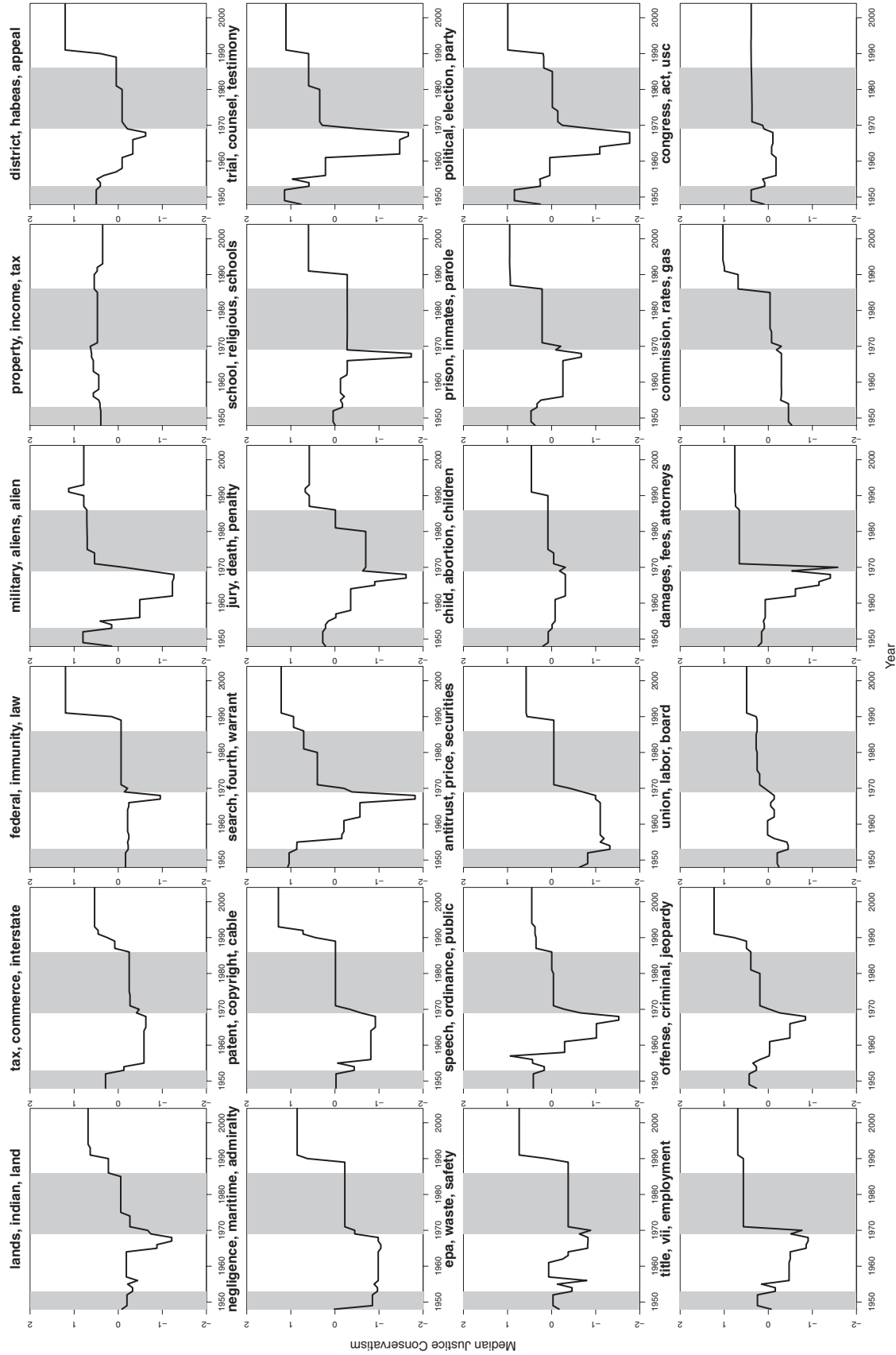
issue-specific consequences of justice replacement are entirely masked by unidimensional preference estimates. For example, most models of judicial preferences suggest the Court became increasingly liberal during the Warren Court and has turned in a conservative direction since, primarily due to justice replacement. This general pattern is reflected in our estimates—the median justice moves to the left on most topics during the Warren Court. However, that overall trend has a number of interesting exceptions, which demonstrate the heterogeneity of preferences across topics. Consider the topics associated with economic issues, such as "antitrust," "price," and "securities"; "union," "labor," and "board"; and "commission," "rates," and "gas." These topics are associated with little change, or even conservative change, during the Warren Court. Indeed, the justices appointed by President Franklin Roosevelt, who served until the 1950s, were particularly liberal on topics of economic regulation (they were chosen for that very reason), and as a consequence, their replacements during the Warren Court actually shifted the Court to the right, rather than the left, on these topics. At the same time, we find the anticipated pattern that the Warren Court moved farthest to the left on issues relating to civil rights and civil liberties.

The location of the median justice across the various topics similarly helps identify when, and on what issues, the Court has moved in a conservative direction since the late 1960s. A conventional understanding of the Burger Court is that President Richard Nixon's attempt to move the Court to the right was a failure, though some have recently argued that this is largely because Nixon was focused on a few issues and did not really want to effect a wholesale conservative revolution on the bench (e.g., McMahon 2011). Figure 6 reveals that the Nixon appointees in fact effected a fairly dramatic conservative shift on issues of civil liberties and criminal procedure, whereas they brought about a much smaller shift on other issues, such as labor rights and employment discrimination, economic regulation, and religion.

This discussion of topic-by-topic preferences requires some caveats. Only a few of the cases are assigned entirely to a single topic; many cases are a mix of topics. While this motivates the structure of the model, it also raises the question of how we should interpret preferences on a single dimension. One answer is to note that the distribution of posterior assignment probabilities is highly bimodal, with posterior probabilities nearly 0% for most topics and over 90% for a single topic. Thus, many cases map mostly to a single topic. The notable exception to this pattern is the topic associated with procedural issues and statutory interpretation ("congress," "act," "usc"). Some cases are entirely about procedural issues, but among those that

⁵The recovered dimensions are correlated, but a secondary analysis of the dimension-specific ideal points using principle components analysis indicated that the variation across justices beyond average left-right position is not explained by one or two additional dimensions: it is issue-specific variation. We think future analyses, especially exploratory roll-call analyses, may benefit from more extensive secondary analyses along these lines.

FIGURE 6 Location of median justice over time for each issue.



Note: Each panel shows the location of the median justice for a given topic, over time. The title of each panel is the top three most distinctive terms for the issue. Medians are calculated term-by-term; terms with Court composition change use the longest Court composition during that term.

are not, there is often a considerable procedural question presented in the case. A second answer is that even in the absence of a case that is entirely contained within a single topic, knowing the Court's median with respect to that topic informs our understanding of how cases implicating the topic change over time. For example, at the same time that the median justice was becoming increasingly liberal on issues of free speech and criminal procedure (during the Warren Court), the median was not changing (or even becoming more conservative) on issues of religion and public schools and labor-business relations. The implication is that cases blending free speech and union issues would experience a less liberal shift than a case blending free speech and criminal procedure. The issue-specific ideal points can be thought of as indices of ideal types: measuring relative conservatism within a specific issue, recognizing that many cases will blend together multiple topics.

Implications

The measurement model we have proposed does more than simply introduce a substantively grounded multidimensional ideal point model that is portable to any number of institutional settings. It also opens the door to new theoretical questions scholars were not previously able to study. In this section, we briefly describe three of the applications and implications of our estimates of Supreme Court ideal points: the politics of judicial nominations, the development of law, and roll-call analysis.

The Politics of Judicial Replacement

Students of politics have spent considerable time studying the impact of executive nominations to courts, agencies, and other bodies. Confirmation battles over Supreme Court nominees have become particularly contentious over the past 30 years. In this vein, the literature has increasingly focused on the extent to which any given nomination will affect the median of the Court, as that is usually seen as an important indicator of the policy consequences of a nomination (Krehbiel 2007). Our analysis demonstrates that the consequences of replacing a justice may vary across areas of the law because the justices differ in their relative rank orderings across areas of the law (see Lauderdale and Clark 2012).

As an illustrative example of how an analysis of confirmation politics may play out, consider the replacement

TABLE 2 Comparison of Disposition Directions on Eight Selected Issues before and after O'Connor's Departure from the Bench

	1994–2004	2005–2010
<i>Kennedy More Conservative</i>		
Sex discrimination in employment	6/7 (86%)	3/5 (60%)
Abortion	3/7 (43%)	1/3 (33%)
Death penalty	5/12 (42%)	3/12 (25%)
Establishment of religion	3/5 (60%)	0/1 (0%)
<i>Kennedy More Liberal</i>		
Securities regulation	1/1 (100%)	5/5 (100%)
Employment arbitration	2/5 (40%)	1/1 (100%)
Patents and copyrights	12/18 (66%)	8/10 (80%)
Attorneys' fees	1/5 (20%)	3/9 (33%)

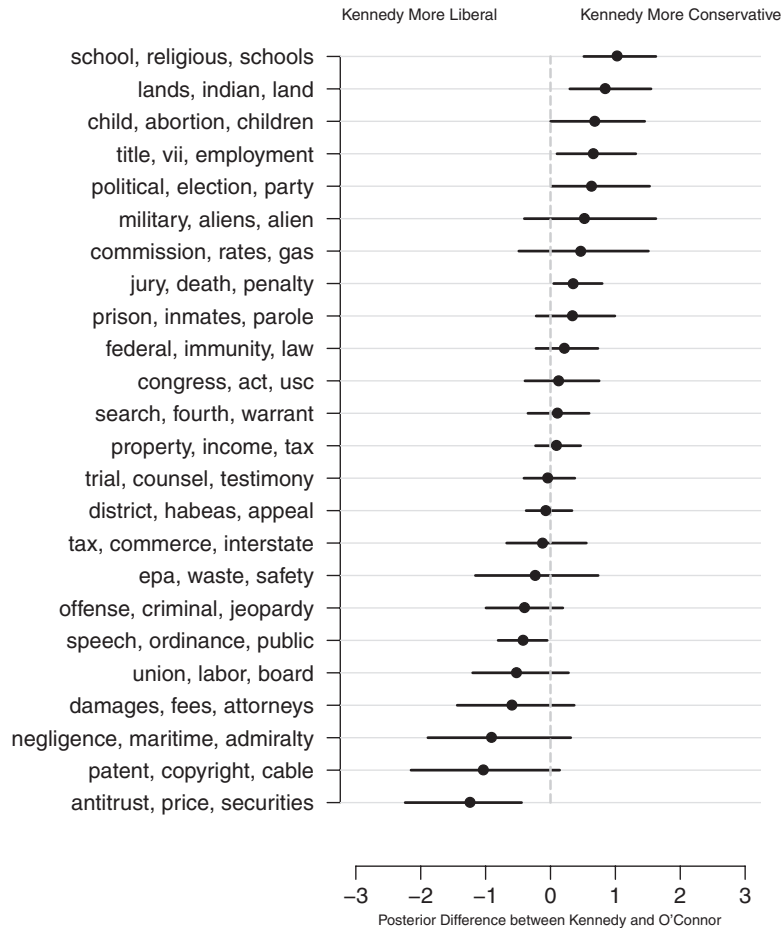
Notes: The table compares the direction of the case disposition on eight issues during the long Rehnquist Court (October terms, 1994–2005) with the Roberts Court (October terms, 2005–2010). Entries are the proportion of decisions that are coded as having liberal dispositions in the Supreme Court Database.

of Justice O'Connor by Justice Alito in 2005. In general, through the 1990s and early 2000s, Justice O'Connor and Justice Kennedy were two pivotal members of the Court. With four relatively liberal justices (Stevens, Souter, Ginsburg, and Breyer) and three relatively conservative justices (Rehnquist, Scalia, and Thomas), how these two justices voted on any given case was often dispositive.

Figure 7 shows the posterior distribution of the difference between Justice Kennedy and Justice O'Connor's positions for each of our 24 dimensions. Positive values indicate Kennedy being more conservative than O'Connor; negative values indicate Kennedy being more liberal than O'Connor. The plots reveal that we have high posterior confidence that O'Connor was to the left of Kennedy on roughly six issues, whereas Kennedy was to the left of O'Connor on roughly four issues. Importantly, this comparison helps illuminate the substantive impact of O'Connor's departure from the Court in 2005, which is widely believed to have left Kennedy as the dominant median justice on the bench (Lane 2006; Lithwick 2006). However, our estimates suggest the ideological impact of this change should depend on the subject matter.

Indeed, these expectations are borne out in the data. Table 2 compares the ideological orientation of the Court's disposition—as coded in the Supreme Court Database (Spaeth et al. n.d.)—in cases involving eight issues both before and after O'Connor's departure from the Court. In the top half of the table, we consider the cases

FIGURE 7 Comparison of Justice Kennedy’s and Justice O’Connor’s Ideal Points on Each Issue dimension



Note: Each row shows the posterior distribution of the difference between Kennedy’s and O’Connor’s ideal points, with the median posterior and central 90% interval depicted along the x-axis. The label of each row is the top three most distinctive terms for the issue.

where our preference estimates suggest her replacement by Alito should have a conservative effect—employment discrimination, abortion, the death penalty, and religious freedom. While the numbers of cases are small, we see that in each of these areas, the rate of liberal decisions has declined. In the bottom half of the table, we consider the cases where our preference estimates suggest her replacement by Alito should have little effect—securities regulation, employment arbitration, patents and copyrights, and attorneys’ fees. In these four areas, we see no shift toward conservative decisions; in fact, if anything, there is a shift in the liberal direction since O’Connor left the Court.

In a small voting chamber like the U.S. Supreme Court, variation in preferences across issues can be highly consequential. Consider the landmark ruling in *Citi-*

zens United v. FEC (2010), which held corporations may spend unlimited amounts of money on behalf of political campaigns. The ruling has had significant implications for American politics in the years since, and there is good reason to believe that the 5–4 decision would have come out the other way had Sandra Day O’Connor still been on the Court. Recent journalistic accounts note that in earlier challenges to the campaign finance restrictions, O’Connor had sided in support of the FEC, and it was only after her replacement by Justice Alito that the Court tilted in favor of corporate spending (Toobin 2012). As we see in Figure 7, Kennedy is indeed to the right of O’Connor on the “political,” “election,” “party” topic.

The upshot of this discussion is that we might expect different kinds of struggles to emerge during a judicial

nomination, as a justice's departure may affect different issues differently. When Kennedy departs the Court, there will likely be much debate, as he is currently so pivotal on many issues. While our approach cannot provide estimates by issue for a nominee before he or she joins the Court, it can identify those areas of law in which Kennedy is the median, and therefore his replacement is guaranteed to move the median in one direction or the other. In areas where Kennedy is not the median, a new justice will only shift the median of the Court if the new justice is on the opposite side of the median from Kennedy.

The Development of Law

A second implication and potential future application of our analysis concerns the development of law in the U.S. Supreme Court. Scholars have long been interested in how doctrine and case law build sequentially through a series of decisions, and much attention has been paid to the use of strategic litigation to shape development of the law. Some of that scholarship has focused on the incentives created by changing membership on the Court to being new cases. Our analysis suggests that, even given a fixed membership on the Court, a strategic litigant may be able to shape the law by framing cases in a particular way. For example, our analysis reveals that Justice Kennedy is relatively conservative on topics of religion and relatively liberal on topics of individual freedom.

Theoretical models of multidimensional preferences in judicial rule making have gained more attention in recent years (Lax 2007). One of the central questions in the economic and political science analysis of legal doctrine concerns how different cases, presenting different mixes of issues, interact over time to build a body of doctrine (Gennaioli and Shleifer 2007; Kornhauser 1992). Indeed, some empirical work has been directly concerned with this dynamic (e.g., Epstein and Kobylka 1992; Wedeking 2010). Our methodology provides an opportunity to gain empirical leverage on the theoretical tensions those models isolate. Our approach allows analysts to investigate the dimensionality of judicial voting and evaluate theoretical predictions about how different mixes of dimensions will split judges differently. This type of analysis might be particularly useful for the analysis of strategic litigation by policy-minded litigants in combination with methods for scaling the content of the opinions themselves (e.g., Clark and Lauderdale 2010).

Roll-Call Analysis

A third implication and potential future application of our methodology extends beyond the judicial politics examples on which we have focused thus far. The method we propose for discovering and describing the dimensionality of a voting space is particularly attractive for estimating multidimensional ideal points in small chambers. However, even in the context of a large voting body, where existing methods work more reliably, our method has the advantage that the dimensions are substantively defined. The requirements for any future application are only (1) a matrix of voting data and (2) a source of metadata that can be used to generate mixture or discrete assignments to dimensions. Here we have used the texts of judicial opinions to assign cases to a set of dimensions. Alternatively, one might use expert codings of case topics (such as the issue codes in the U.S. Supreme Court Database or the topic codes from the Policy Agendas Project). The critical component is that one is able to generate a matrix $\mathbf{\Lambda}$ that defines a meaningful political space in which to map voting behavior.

The rise of widely available metadata on political voting—such as text of debates and legislation—is, we believe, a particularly promising avenue for such applications in the future. Students of supranational legislative and judicial institutions, for example, have access to increasingly extensive textual information and also are becoming increasingly interested in more nuanced descriptions of cleavages on the Inter-American Court of Human Rights, the European Court of Human Rights, the European Parliament, the United Nations General Assembly, and the World Trade Organization, among others. One of the most compelling applications of our method is exploratory: it can be used in conjunction with many forms of metadata to discover and summarize the nature of political cleavages in both large and small voting bodies.

Conclusion

The model we describe in this article does not generate sets of ideal point estimates that we can call *the* ideal points of the actors. There is no true number of issues to be uncovered; the number of issues we can usefully estimate is in part a function of data density, rather than any fundamental truth about how many issues a voting body considers. Indeed, in our example, a wide range of values for the number of estimated topics all yield roughly equal values of the DIC, the criterion that we use to evaluate fit. However, this limitation is fundamental to all

scaling models. Ideal point estimates exist on scales that researchers create: they are useful fictions that help us describe patterns in large data sets. Similarly, our approach does not remove the need to carefully consider whether a scale is appropriate to use as an explanatory variable in a secondary analysis (Martin and Quinn n.d.). Analysts need to consider whether using measures of issue-specific ideology that are derived from voting behavior—whether at large or in a domain identified by word use in opinion (or other) texts—creates a threat to the kind of inference that they intend to make.

Unidimensional ideal points provide especially parsimonious summaries of political conflict in legislative and judicial bodies, reducing a large matrix of voting patterns to a single measure of relative preference. But while these are useful first-order approximations to behavior, politics is not so simple as one (or even two) dimension(s). Methodological obstacles have kept scholars from getting much more from ideal point estimators than the first two dimensions. Even in two dimensions, identification and interpretation problems quickly dominate the process of analyzing roll call data: only by simultaneously examining actors with extreme positions and votes with large loadings on particular dimensions can labels be put on a standard ideal point estimates, whether those estimates come from NOMINATE, optimal classification or Bayesian IRT. Beyond two dimensions, these informal labeling techniques quickly become infeasible. Researchers have sought to overcome these problems by leveraging metadata to overcome identification and interpretation problems (e.g., Clinton and Meirowitz 2003; Jessee 2009; Lauderdale and Clark 2012; Zucco and Lauderdale 2011), though these approaches have tended to be methodologically tied to the particular institutions under study in ways that limit their general applicability.

Our approach provides better descriptions than previous ideal point estimation techniques because it addresses the reductionism problem and the labeling problem simultaneously, while also being generally applicable to the full variety of roll-call data that researchers have examined. Using textual data enables more nuanced descriptions of voting behavior, with that nuance provided in terms of the actual terms that describe the issues at stake. Unsupervised text models like LDA, applied to vote-specific texts, are an excellent tool for exploratory analyses like the one we present here. Supervised text models, which impose a predefined categorization, will be useful for estimating issue preferences on particular topics of interest to researchers. More broadly, any clustering algorithm or mixture model, applied to vote-specific metadata that vary primarily by the topic of the vote, will produce a valid λ matrix of issue weights for our mod-

ified IRT model. For example, if one had network data describing which votes were on similar issues (Lauderdale and Clark 2012), one could use community detection methods (Porter, Onnela, and Mucha 2009) to generate a suitable discrete categorization. Given the range of metadata on votes that exists across institutions, our approach is potentially applicable to most existing roll-call data sets: any voting data for which there are a set of substantively relevant documents associated with the votes, such as recorded debates, bills, court opinions, treaties, and resolutions. This methodology will be especially useful for the study of small voting bodies like the Supreme Court, where estimating typical multidimensional models is difficult because there are not enough observations for each vote. Our approach will also be especially useful for highly multidimensional voting bodies like the European Parliament (Hix, Noury, and Roland 2006). In a voting body like the U.S. Congress, there is less variation to describe beyond the first dimension (Poole and Rosenthal 1997), but even in Congress there are some legislators who are not fit well by a unidimensional model because their preferences vary across different issues (Lauderdale 2010).

In our analysis of the U.S. Supreme Court, we have documented not only the extent to which different subjects have dominated the Court's docket over time, but also how the justices' preferences have systematically varied across those issues. Political scientists have collected a wide range of roll-call data over the past two decades and characterized one or two aggregate dimensions of conflict in many voting bodies. By leveraging metadata about votes and the increasing variety of classification techniques, we can more richly describe political cleavages.

References

- Blei, David, and John D. Lafferty. 2009. "Topic Models." In *Text Mining: Classification, Clustering, and Applications*, ed. A. Srivastava, and M. Sahami. Data Mining and Knowledge Discovery. Taylor & Francis, London, England, 2009.
- Blei, David M., and Jon D. McAuliffe. 2007. "Supervised Topic Models." *Neural Information Processing Systems* 21: 1280–1287.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3: 993–1022.
- Clark, Tom S., and Benjamin Lauderdale. 2010. "Locating Supreme Court Opinions in Doctrine Space." *American Journal of Political Science* 54(4): 871–90.
- Clark, Tom S., and Benjamin Lauderdale. 2012. "The Genealogy of Law." *Political Analysis* 20(3): 329–50.

- Clinton, Joshua D., Simon Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98(2):355–70.
- Clinton, Joshua D., and Adam Meirowitz. 2003. "Integrating Roll Call Analysis and Voting Theory: A Framework." *Political Analysis* 11: 381–96.
- Crespin, Michael H., and David W. Rohde. 2010. "Dimensions, Issues, and Bills: Appropriations Voting on the House Floor." *Journal of Politics* 72(4): 976–89.
- Epstein, Lee, and Joseph F. Kobylka. 1992. *The Supreme Court and Legal Change*. Chapel Hill: University of North Carolina Press.
- Gennaioli, Nicola, and Andrei Shleifer. 2007. "The Evolution of Common Law." *Journal of Political Economy* 115(1): 43–68.
- Griffiths, Thomas L., and Mark Steyvers. 2004. "Finding Scientific Topics." *Proceedings of the National Academy of Sciences* 101: 5228–35.
- Grofman, Bernard, and Timothy J. Brazill. 2002. "Identifying the Median Justice on the Supreme Court through Multidimensional Scaling: Analysis of 'Natural Courts' 1953–1991." *Public Choice* 112(1/2): 55–79.
- Heinrich, Gregor. 2009. "Parameter Estimation for Text Analysis." Technical Report 2.9. Darmstadt, Germany: Fraunhofer IGD.
- Hix, Simon, and Christophe Crombez. 2005. "Extracting Ideal Point Estimates from Actors' Preferences in the EU Constitutional Negotiations." *European Union Politics* 6(3): 353–76.
- Hix, Simon, Abdul Noury, and Gerard Roland. 2006. "Dimensions of Politics in the European Parliament." *American Journal of Political Science* 50(2): 494–511.
- Hopkins, Daniel J., and Gary King. 2010. "A Method of Automated Nonparametric Content Analysis for Social Science." *American Journal of Political Science* 54(1): 229–47.
- Jackman, Simon. 2001. "Multidimensional Analysis of Roll Call Data via Bayesian Simulation: Identification, Estimation, Inference and Model Checking." *Political Analysis* 9(3): 227–41.
- Jeffries, John C., Jr. 1994. *Justice Lewis F. Powell, Jr.: A Biography*. New York: Fordham University Press.
- Jessee, Stephen. 2009. "Spatial Voting in the 2004 Presidential Election." *American Political Science Review* 103: 59–81.
- Kornhauser, Lewis A. 1992. "Modeling Collegial Courts I: Path Dependence." *International Review of Law and Economics* 12: 169–85.
- Krehbiel, Keith. 2007. "Supreme Court Appointments as a Move-the-Median Game." *American Journal of Political Science* 51(2): 231–40.
- Lane, Charles. 2006. "Kennedy Reigns Supreme on Court." *Washington Post*, July 2.
- Lauderdale, Benjamin E. 2010. "Unpredictable Voters in Ideal Point Estimation." *Political Analysis* 18(2): 151–71.
- Lauderdale, Benjamin E., and Tom S. Clark. 2012. "The Supreme Court's Many Median Justices." *American Political Science Review* 106(4): 847–66.
- Lax, Jeffrey R. 2007. "Constructing Legal Rules on Appellate Courts." *American Political Science Review* 101(3): 591–604.
- Lithwick, Dahlia. 2006. "Swing Time." <http://www.slate.com/id/2134421/> January 17, 2006.
- Martin, Andrew D., and Kevin M. Quinn. n.d. "Can Ideal Point Estimates Be Used as Explanatory Variables?" Unpublished manuscript, Washington University in St. Louis, St. Louis, Missouri.
- McMahon, Kevin J. 2011. *Nixon's Court*. Chicago: University of Chicago Press.
- Newman, Roger K. 1997. *Hugo Black: A Biography*. New York: Fordham University Press.
- Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.
- Porter, Mason A., Jukka-Pekka Onnela, and Peter J. Mucha. 2009. "Communities in Networks." *Notices of the American Mathematical Society* 56: 1082–97, 1164–66.
- Rivers, Douglas. 2003. "Identification of Multidimensional Spatial Voting Models." Unpublished manuscript. Stanford University.
- Spaeth, Harold, Lee Epstein, Ted Ruger, Keith Whittington, Jeffrey Segal, and Andrew D. Martin. n.d. "The Supreme Court Database." <http://www.supremecourtdatabase.org>
- Spiegelhalter, David J., Nicola G. Best, Bradley P. Carlin, and Angelika Van der Linde. 2002. "Bayesian Measures of Model Complexity and Fit (with Discussion)." *Journal of the Royal Statistical Society Series B* 64(4): 583–616.
- Stern, Seth, and Stephen Wermiel. 2010. *Justice Brennan: Liberal Champion*. Boston: Houghton Mifflin Harcourt.
- Toobin, Jeffrey. 2012. "Money Unlimited." *The New Yorker*, May 21, 2012 http://www.newyorker.com/reporting/2012/05/21/120521fa_fact_toobin
- Voeten, Erik. 2000. "Clashes in the Assembly." *International Organization* 54(2): 185–215.
- Walker, Thomas G. 2009. *Eligible for Execution: The Story of the Daryl Atkins Case*. Washington, DC: CQ Press.
- Wallach, Hanna M., David Mimno, and Andrew McCallum. 2009. "Rethinking LDA: Why Priors Matter." *Neural Information Processing Systems* 23: 1973–1981.
- Wedeking, Justin. 2010. "Supreme Court Litigants and Strategic Framing." *American Journal of Political Science* 54(3): 617–31.
- Zucco, Cesar, Jr., and Benjamin E. Lauderdale. 2011. "Distinguishing between Influences on Brazilian Legislative Behavior." *Legislative Studies Quarterly* 71(3): 1076–1092.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

These materials include all raw data, code for replication, and estimates described in the paper.