How They Vote: Issue-Adjusted Models of Legislative Behavior

Sean M. Gerrish*
Department of Computer Science
Princeton University
Princeton, NJ 08540
sgerrish@cs.princeton.edu

David M. Blei
Department of Computer Science
Princeton University
Princeton, NJ 08540
blei@cs.princeton.edu

Abstract

We develop a probabilistic model of legislative data that uses the text of the bills to uncover lawmakers' positions on specific political issues. Our model can be used to explore how a lawmaker's voting patterns deviate from what is expected and how that deviation depends on what is being voted on. We derive approximate posterior inference algorithms based on variational methods. Across 12 years of legislative data, we demonstrate both improvement in heldout predictive performance and the model's utility in interpreting an inherently multi-dimensional space.

1 Introduction

Legislative behavior centers around the votes made by lawmakers. Capturing regularity in these votes, and characterizing patterns of legislative behavior, is one of the main goals of quantitative political science. Voting behavior exhibits enough regularity that simple statistical models, particularly *ideal point models*, easily capture the broad political structure of legislative bodies. However, some lawmakers do not fit neatly into the assumptions made by these models. In this paper, we develop a new model of legislative behavior that captures when and how lawmakers vote differently than expected.

Ideal point models assume that lawmakers and bills are represented as points in a latent space. A lawmaker's (stochastic) voting behavior is characterized by the relationship between her position in this space and the bill's position [1, 2, 3]. Given the data of how each lawmaker votes on each bill (known as a roll call), we can use ideal point models to infer the latent position of each lawmaker. In U.S. politics, these inferred positions reveal the commonly-known political spectrum: right-wing lawmakers are at one extreme, and left-wing lawmakers are at the other. Figure 1 illustrates example inferences from an ideal point model.

But there are some votes that ideal point models fail to capture. For example, Ronald Paul, Republican representative from Texas, and Dennis Kucinich, Democratic representative from Ohio, are poorly modeled by ideal points because they diverge from the left-right spectrum on issues like foreign policy. Because some lawmakers deviate from their party on certain issues, their positions on these issues are not captured by ideal point models.

To this end, we develop the *issue-adjusted ideal point model*, a latent variable model of roll-call data that accounts for the contents of the bills that lawmakers are voting on. The idea is that each lawmaker has both a general position and a sparse set of position adjustments, one for each issue. The votes on a bill depend on a lawmaker's position, *adjusted* for the bill's content. The text of the bill encodes the issues it discusses. Our model can be used as an exploratory tool for identifying

*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

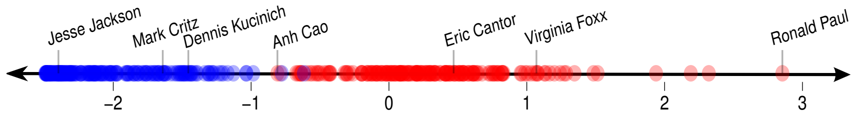


Figure 1: Traditional ideal points separate Republicans (red) from Democrats (blue).

exceptional voting patterns of individual legislators, and it provides a richer description of lawmakers' voting behavior than the models traditionally used in political science.

In the following sections, we develop our model and describe an approximate posterior inference algorithm based on variational methods. We analyze six Congresses (12 years) of legislative data from the United States Congress. We show that our model gives a better fit to legislative data and provides an interesting exploratory tool for analyzing legislative behavior.

2 Exceptional issue voting

We first review ideal point models of legislative roll call data and discuss their limitations. We then present a model that accounts for how legislators vote on specific issues.

Modeling politics with ideal points.

Ideal point models are based on item response theory, a statistical theory that models how members of a population judge a set of items. Applied to voting records, one-dimensional ideal point models place lawmakers on an interpretable political spectrum. These models are widely used in quantitative political science [3, 4, 5].

One-dimensional ideal point models posit an *ideal point* $x_u \in \mathbb{R}$ for each lawmaker u. Each bill d is characterized by its *polarity* a_d and its *popularity* b_d.[1] The probability that lawmaker u votes "Yes" on bill d is given by the logistic regression

$$p(v_{ud} = \text{yes} \mid x_u, a_d, b_d) = \sigma(x_u a_d + b_d), \tag{1}$$

where $\sigma(s) = \frac{\exp(s)}{1+\exp(s)}$ is the logistic function.[2] When the popularity of a bill b_d is high, nearly everyone votes "Yes"; when the popularity is low, nearly everyone votes "No". When the popularity is near zero, the probability that a lawmaker votes "Yes" depends on how her ideal point x_u interacts with bill polarity a_d. The variables a_d, b_d, and x_u are usually assigned standard normal priors [3].

Given a matrix of votes, we can infer the ideal point of each lawmaker. We illustrate ideal points fit to votes in the U.S. House of Representatives from 2009-2010 in Figure 1. The model has clearly separated lawmakers by their political party (colour) and provides an intuitive measure of their political leanings.

Limitations of ideal point models. A one-dimensional ideal point model fit to the U.S. House from 2009-2010 correctly models 98% of lawmakers' votes on training data. But it only captures 83% of Baron Hill's (D-IN) votes and 80% of Ronald Paul's (R-TX) votes. Why is this?

The ideal point model assumes that lawmakers are ordered. Each bill d splits them at a *cut point* $-\frac{b_d}{a_d}$. Lawmakers to one side of the cut point are more likely to support the bill, and lawmakers to the other side are likely to reject it. For lawmakers like Paul and Hill, this assumption is too strong because their voting behavior does not fit neatly into a single ordering. Their location among the other lawmakers changes with different bills.

Lawmakers do not vote randomly, however. They vote consistently within individual areas of policy, such as foreign policy and education. For example, Rep. Paul consistently votes against United States involvement in foreign military engagements, a position that contrasts with other Republicans.

We refer to voting behavior like this as *issue voting*. An *issue* is any federal policy area, such as "financial regulation," "foreign policy," "civil liberties," or "education," on which lawmakers are expected to take positions. Lawmakers' positions on these issues often diverge from their traditional left/right stances. The model we will develop captures these deviations. Some examples are illustrated

[1] These are sometimes called the *discrimination* and *difficulty*, respectively.

[2] Many ideal point models use a probit function instead [1, 3].

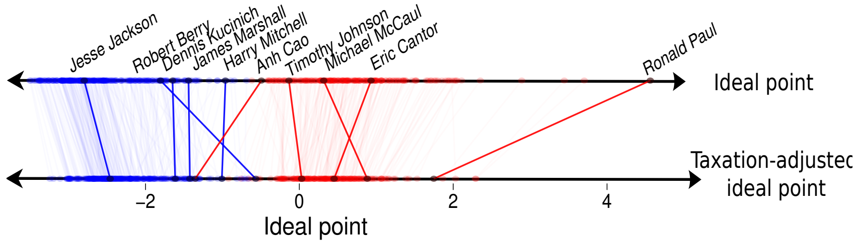
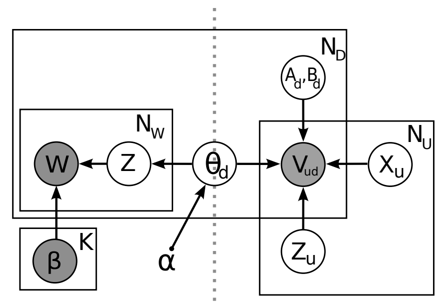


Figure 2: In a traditional ideal point model, lawmakers' ideal points are static (top line of each figure). In the issue-adjusted ideal point model, lawmakers' ideal points change when they vote on certain issues, such as *Taxation*.

Terrorism	Commemorations	Transportation
terrorist	nation	transportation
september	people	minor
attack	life	print
nation	world	tax
york	serve	land
terrorist attack	percent	guard
hezbolah	community	coast guard
national guard	family	substitute

Labeled topics



The issue-adjusted ideal point model

Figure 3: Left: Top words from topics fit using labeled LDA [6]. Right: the issue-adjusted ideal point model, which models votes v_{ud} from lawmakers and legislative items. Classic item response theory models votes v using x_u and a_d, b_d. For our work, documents' issue vectors $\boldsymbol{\theta}$ were estimated fit with a topic model (left of dashed line) using bills' words w and labeled topics β. Expected issue vectors $\mathbb{E}_q [\boldsymbol{\theta}|\boldsymbol{w}]$ are then treated as constants in the issue model (right of dashed line).

in Figure 2; Charles Djou is more similar to Republicans on *Taxation* (right) and more similar to Democrats on *Health* (left), while Ronald Paul is more Republican-leaning on *Health* and less extreme on *Taxation*. The model we will introduce uses lawmakers' votes and the text of bills to model deviations like this, on a variety of issues. This allows us to take into account whether a bill was about *Taxation* or *Education* (or both) when predicting a lawmaker's vote.

Issue-adjusted ideal points.

We now describe the issue-adjusted ideal point model, a new model of lawmaker behavior that takes into account both the content of the bills and the voting patterns of the lawmakers. We build on the ideal point model so that each lawmaker's ideal point can be adjusted for each issue.

Suppose that there are K issues in the political landscape. We will use the words $\boldsymbol{w}_d$ of each bill d to code it with a *mixture* $\boldsymbol{\theta}_d$ of issues, where each element θ_{dk} corresponds to an issue; the components of $\boldsymbol{\theta}_d$ are positive and sum to one. (These vectors will come from a topic model, which we describe below.) In our proposed model, each lawmaker is also associated with a K-vector $\boldsymbol{z}_u \in \mathbb{R}^K$, which describes how her ideal point changes for bills about each issue.

We use these variables in a model based on the traditional ideal point model of Equation 1. As above, x_u is the ideal point for lawmaker u and a_d, b_d are the polarity and popularity of bill d. In our model, votes are modeled with a logistic regression

$$p(v_{ud}|a_d, b_d, z_u, x_u, \boldsymbol{w}_d) = \sigma\left((\boldsymbol{z}_u^\top \mathbb{E}_q [\boldsymbol{\theta}_d|\boldsymbol{w}_d] + x_u)a_d + b_d\right), \tag{2}$$

where we use an estimate $\mathbb{E}_q [\boldsymbol{\theta}_d|\boldsymbol{w}_d]$ of the bill's issue vector from its words $\boldsymbol{w}_d$ as described below.

We put standard normal priors on the ideal points, polarity, and difficulty variables. We use Laplace priors for $\boldsymbol{z}_u$: $p(z_{uk} \mid \lambda_1) \propto \exp(-\lambda_1 ||z_{uk}||_1)$. This enforces a sparse penalty with MAP inference and a "nearly-sparse" penalty with Bayesian inference. See Figure 3 (left) for the graphical model.

To better understand the model, assume that bill d is only about *Finance*. This means that $\boldsymbol{\theta}_d$ has a one in the *Finance* dimension and zero everywhere else. With a classic ideal point model, a lawmaker u's ideal point, x_u, gives his position on each issue, including *Finance*. With the issue-adjusted ideal point model, his *effective ideal point* for *Finance*, $x_u + z_{u,\text{Finance}}$, gives his position on *Finance*. The adjustment $z_{u,\text{Finance}}$ affects how lawmaker u feels about *Finance* alone. When $z_{u,k} = 0$ for all u, k, the model becomes the classic ideal point model.

This model lets us inspect lawmakers' overall voting patterns by issue. Given a collection of votes and a coding of bills to issues, posterior estimates of the ideal points and per-issue adjustments give us a window into voting behavior that is not available to classic ideal point models.

Using Labeled LDA to associate bills with issues.

Equation 2 adjusts a lawmaker's ideal point by using the conditional expectation of a bill's thematic labels $\boldsymbol{\theta}_d$ given its words $\boldsymbol{w}_d$. We estimate this vector using labeled latent Dirichlet allocation (LDA) [6]. Labeled LDA is a topic model, a bag-of-words model that assumes a set of themes for the collection of bills and that each bill exhibits a mixture of those themes. The themes, called topics, are distributions over a fixed vocabulary. In unsupervised LDA [7] they are learned from the data. In labeled LDA, they are defined by using an existing tagging scheme. Each tag is associated with a topic; its distribution is found by taking the empirical distribution of words for documents assigned to that tag.[3] This gives interpretable names (the tags) to the topics.

We used tags provided by the Congressional Research Service [8], which provides subject codes for all bills passing through Congress. These subject codes describe the bills using phrases which correspond to traditional issues, such as *Civil rights* and *National security*. Each bill may cover multiple issues, so multiple codes may apply to each bill. (Many bills have more than twenty labels.) We used the 74 most-frequent issue labels. Figure 3 (right) illustrates the top words from several of these labeled topics.[4] We fit the issue vectors $\mathbb{E}\left[\boldsymbol{\theta}_d | \boldsymbol{w}_d\right]$ as a preprocessing step. In the issue-adjusted ideal point model (Equation 2), $\mathbb{E}\left[\boldsymbol{\theta}_d\right]$ was treated as observed when estimating the posterior distribution $p(x_u, a_d, b_d, \boldsymbol{z}_d | \mathbb{E}\left[\boldsymbol{\theta}_d | \boldsymbol{w}_d\right], v_{ud})$. We summarize all 74 issue labels in §A.2.[5]

Related Work. Item response theory has been used for decades in political science [3, 4, 5]; see Enelow and Hinich for a historical perspective [9] and Albert for Bayesian treatments of the model [10]. Some political scientists have used higher-dimensional ideal points, where each legislator is attached to a vector of ideal points $\boldsymbol{x}_u \in \mathbb{R}^K$ and each bill polarization $\boldsymbol{a}_d$ takes the same dimension K [11]. The probability of a lawmaker voting "Yes" is $\sigma(\boldsymbol{x}_u^T \boldsymbol{a}_d + b_d)$. The principal component of ideal points explains most of the variance and explains party affiliation. However, other dimensions are not attached to issues, and interpreting beyond the principal component is painstaking [2].

Recent work in machine learning has provided joint models of legislative text and the bill-making process. This includes using transcripts of U.S. Congressional floor debates to predict whether speeches support or oppose pending legislation [12] and predicting whether a bill will survive congressional committee by incorporating a number of features, including bill text [13]. Other work has aimed to predict individual votes. Gerrish and Blei aimed to predict votes on bills which had not yet received any votes [14]. Their model fits a_d and b_d using supervised topics, but the underlying voting model was one-dimensional: it could not model individual votes better than a one-dimensional ideal point model. Wang et al. created a Bayesian nonparametric model of votes and text over time [15]. We note that these models have different purposes from ours, and neither addresses individuals' affinity toward issues.

The issue-adjusted model is conceptually more similar to recent models for content recommendation. Wang and Blei describe a method to recommend academic articles to individuals [16], and Agarwal and Chen propose a model to match users to Web content [17]. Though they do not consider roll-call data, these recommendation models also try to match user behavior with textual item content.

[3] Ramage et al. explore more sophisticated approaches [6], but we found this simplified version to work well.

[4] After defining topics, we performed two iterations of LDA with variational inference to smooth the topics.

[5] We refer to specific sections in the supplementary materials (appendix) as §A.#.

3 Posterior estimation

The central computational challenge in this model is to uncover lawmakers' issue preferences $\boldsymbol{z}_u$ by using the their votes v and bills' issues $\boldsymbol{\theta}_d$. We do this by estimating the posterior distribution $p(x, \boldsymbol{z}, a, b|v, \boldsymbol{\theta})$. Bayesian ideal point models are usually fit with Gibbs sampling [2, 3, 5, 18]. However, fast Gibbs samplers are unavailable for our model because the conditionals needed are not analytically computable. We estimate the posterior with variational Bayes.

In variational Bayes, we posit a family of distributions $\{q_{\boldsymbol{\eta}}\}$ over the latent variables that is likely to contain a distribution similar to the true posterior [19]. This variational family is indexed by parameters $\boldsymbol{\eta}$, which are fit to minimize the KL divergence between the variational and true posteriors. Specifically, we let $\{q_{\boldsymbol{\eta}}\}$ be the family of fully factorized distributions

$$q(x, \boldsymbol{z}, a, b|\boldsymbol{\eta}) = \prod_U \mathcal{N}(x_u|\tilde{x}_u, \sigma^2_{x_u})\mathcal{N}(\boldsymbol{z}_u|\tilde{z}_u, \lambda_{z_u}) \prod_D \mathcal{N}(a_d|\tilde{a}_d, \sigma^2_{a_d})\mathcal{N}(b_d|\tilde{b}_d, \sigma^2_{b_d}), \tag{3}$$

where we parameterize the variational posterior with $\boldsymbol{\eta} = \{(\tilde{x}_u, \sigma_x), (\tilde{z}_u, \sigma_{\boldsymbol{z}_u}), (\tilde{a}, \sigma_a), (\tilde{b}, \sigma_b)\}$. We assumed full factorization to make inference tractable. Though simpler than the true posterior, fitted variational distributions can be excellent proxies for it. The similarity between ideal points fit with variational inference and MCMC has been demonstrated in Gerrish in Blei [14].

Variational inference usually proceeds by optimizing the variational objective

$$\mathcal{L}_{\boldsymbol{\eta}} = \mathbb{E}_{q_{\boldsymbol{\eta}}}\left[\log p(x, \boldsymbol{z}, a, b, v, \boldsymbol{\theta})\right] - \mathbb{E}_{q_{\boldsymbol{\eta}}}\left[\log q_{\boldsymbol{\eta}}(x, \boldsymbol{z}, a, b)\right] \tag{4}$$

with gradient or coordinate ascent (this is equivalent to optimizing the KL divergence between q and the posterior). Optimizing this bound is challenging when the expectation is not analytical, which makes computing the exact gradient $\nabla_{\boldsymbol{\eta}}\mathcal{L}_{\boldsymbol{\eta}}$ more difficult. We optimize this bound with stochastic gradient ascent [20, 21], approximating the gradient with samples from $q_{\boldsymbol{\eta}}$;

$$\nabla_{\boldsymbol{\eta}}\mathcal{L}_{\boldsymbol{\eta}} \approx \frac{1}{M}\sum_{y_m \sim q_{\boldsymbol{\eta}}} \frac{\partial q_{\boldsymbol{\eta}}}{\partial \boldsymbol{\eta}}(\log p(y_m, v, \boldsymbol{\theta}) - \log q_{\boldsymbol{\eta}}(y_m)); \tag{5}$$

where $y_m = (x_m, \boldsymbol{z}_m, a_m, b_m)$ is a sample from $q_{\boldsymbol{\eta}}$. The algorithm proceeds by following this stochastic gradient with decreasing step size; we provide further details in §A.1.

4 Analyzing twelve years of U.S. legislative history

We used our model to investigate twelve years of U.S. legislative history. We compare the posterior fit with this model to the same data fit with traditional ideal points and validate the model quantitatively. We then provide a closer look at the collection of issues, lawmakers, and bills and explore several interesting results of the model.

4.1 Data and Experiment Setup

We studied U.S. Senate and House of Representative roll-call votes from 1999 to 2010. This period spanned Congresses 106 to 111 and covered an historic period in recent U.S. politics, the majority of which Republican President George W. Bush held office. Bush's inauguration and the attacks of September 11th, 2001 marked the first quarter of this period, followed by the wars in Iraq and Afghanistan. Congress became more partisan over this period, and Democratic President Obama was inaugurated in January 2009.

We provide a more complete summary of statistics for our datasets in §A.3. For context, the median session we considered had 540 lawmakers, 507 bills, and 201,061 votes in both the House and Senate. Altogether, there were 865 unique lawmakers, 3,113 bills, and 1,208,709 votes.

Corpus preparation. For each congress, we considered only bills for which votes were explicitly recorded in a roll-call. We ignored votes on bills for which text was unavailable. To fit the labeled topic model to each bill, we removed stop words and grouped common phrases as n-grams. All bills were downloaded from `www.govtrack.us` [22], a nonpartisan website which provides records of U.S. Congressional voting. We fit the Senate and House separately for each two-year Congress because lawmakers' strategies change at each session boundary.

Table 1: Average log-likelihood of heldout votes using six-fold cross validation. These results cover Congresses 106 to 111 (1999-2010) with regularization $\lambda = 1$. The issue-adjusted model yields higher heldout log-likelihood for all congresses in both chambers than a standard ideal point model. Perm. Issue illustrates the issue model fit when bills' issue labels were randomly permuted. *Perm. Issue* is results for the issue model fit using permuted document labels.

Model	**Senate**					
Congress	106	107	108	109	110	111
Ideal	-0.209	-0.209	-0.182	-0.189	-0.206	-0.182
Issue	**-0.208**	**-0.209**	**-0.181**	**-0.188**	**-0.205**	**-0.180**
Perm. Issue	-0.210	-0.210	-0.183	-0.203	-0.211	-0.186
	House					
Ideal	-0.168	-0.154	-0.096	-0.120	-0.090	-0.182
Issue	**-0.166**	**-0.147**	**-0.093**	**-0.116**	**-0.087**	**-0.180**
Perm. Issue	-0.210	-0.211	-0.100	-0.123	-0.098	-0.187

4.2 Comparison of classic and exploratory ideal points

How do classic ideal points compare with issue-adjusted ideal points? We fit classic ideal points to the 111th House (2009 to 2010) to compare them with issue-adjusted ideal points $\tilde{x}_u$ from the same period, using regularization $\lambda = 1$. The models' ideal points $\tilde{x}_u$ were very similar, correlated at 0.998. While traditional ideal points cleanly separate Democrats and Republicans in this period, issue-adjusted ideal points provide an even cleaner break between the parties. Although the issue-adjusted model is able to use other parameters—lawmakers' adjustments $\tilde{z}_u$—to separate the parties better, the improvement is much greater than expected by chance ($p < 0.001$ using a permutation test).

4.3 Evaluation and significance

We first evaluate the issue-adjusted model by measuring how it can predict held out votes. (This is a measure of model fitness.) We used six fold cross-validation. For each fold, we computed the average predictive log-likelihood $\log p(v_{ud\text{Test}} | v_{ud\text{Train}}) = \log p(v_{ud\text{Test}} | \tilde{x}_u, \tilde{z}_u, \tilde{a}_d, \tilde{b}_d, \mathbb{E}_q [\boldsymbol{\theta}_d | \boldsymbol{w}])$ of the test votes and averaged this across folds. We compared these with the ideal point model, evaluating the latter in the same way. We give implementation details of the model fit in §A.1.

Note that we cannot evaluate how well this model predicts votes on a heldout bill d. As with the ideal point model, our model cannot predict $\tilde{a}_d, \tilde{b}_d$ without votes on d. Gerrish and Blei [14] accomplished this by predicting $\tilde{a}_d$ and $\tilde{b}_d$ using the document's text. (Combining these two models would be straightforward.)

Performance. We compared the issue-adjusted model's ability to represent heldout votes with the ideal point model. We fit the issue-adjusted model to both the House and Senate for Congresses 106 to 110 (1999-2010) with regularization $\lambda = 1$. For comparison we also fit an ideal point model to each of these congresses. In all Congresses and both chambers, the issue-adjusted model represents heldout votes with higher log-likelihood than an ideal point model. We show these results in Table 1.

Sensitivity to regularization. To measure sensitivity to parameters, we fit the issue-adjusted model to the 109th Congress (1999-2000) of the House and Senate for a range $\lambda = 0.0001, \ldots, 1000$ of regularizations. We fixed variance $\sigma_X^2, \sigma_Z^2, \sigma_A^2, \sigma_B^2 = \exp(-5)$. The variational implementation generalized well for the entire range, with heldout log likelihood highest for $1 \le \lambda \le 10$.

Permutation test. We used a permutation test to understand how the issue-adjusted model improves upon ideal point models. This test strengthens the argument that *issues* (and not some other model change, such as the increase in dimension) help to improve predictive performance. To do this test, we randomly permuted topic vectors' document labels to completely remove the relationship between topics and bills: $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_D) \mapsto (\boldsymbol{\theta}_{\pi_i(1)} \ldots \boldsymbol{\theta}_{\pi_i(D)})$, for five permutations $\pi_1, \ldots, \pi_5$. We then fit the issue model using these permuted document labels. As shown in Table 1, models fit with the original, unpermuted issues always formed better predictions than models fit with the permuted issues. From this, we draw the conclusion that *issues* indeed help the model to represent votes.

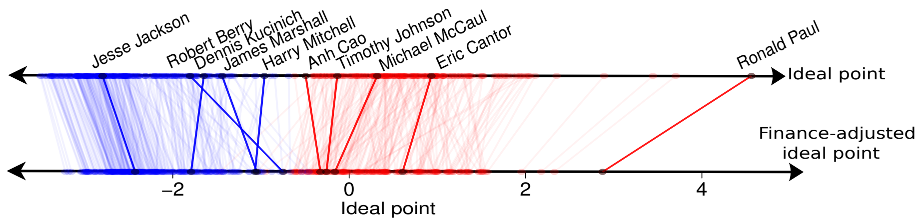


Figure 4: Ideal points x_u and issue-adjusted ideal points $x_u + z_{uk}$ from the 111th House for the *Finance* issue. Republicans (red) saw more adjustment than Democrats (blue).

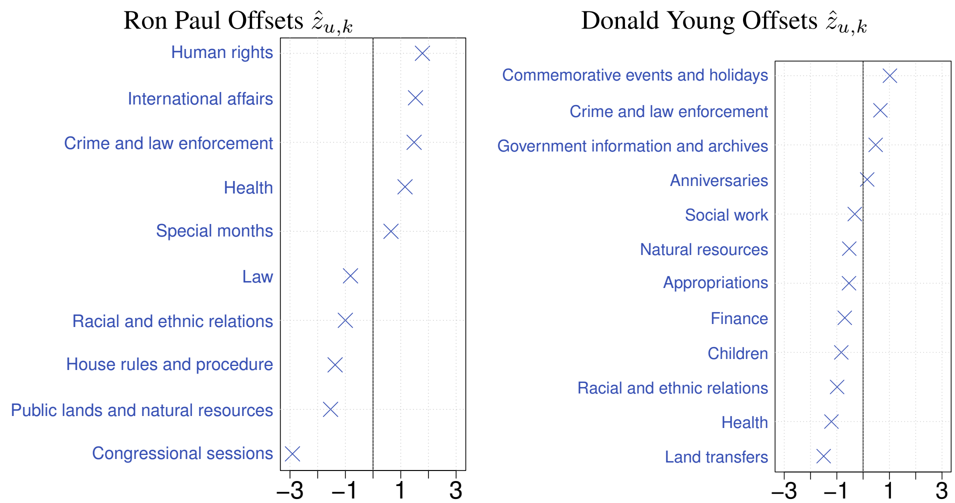


Figure 5: Significant issue adjustments for exceptional senators in Congress 111. Statistically significant issue adjustments are shown with each $\times$.

4.4 Analyzing issues, lawmakers, and bills

In this section we take a closer look at how issue adjustments improve on ideal points and demonstrate how the issue-adjusted ideal point model can be used to analyze specific lawmakers. We focus on an issue-adjusted model fit to all votes in the 111th House of Representatives (2009-2010).

We can measure the improvement by comparing the training likelihoods of votes in the issue-adjusted and traditional ideal point models. The training log-likelihood of each vote is

$$J_{ud} = 1_{\{v_{ud}=\text{Yes}\}} p - \log(1 + \exp(p)), \tag{6}$$

where $p = (\tilde{x}_u + \tilde{z}_u^T \mathbb{E}_q [\boldsymbol{\theta}_d | \boldsymbol{w}])\tilde{a}_d + \tilde{b}_d$ is the log-odds of a vote under the issue adjusted voting model. The corresponding log-likelihood I_{ud} under the ideal point model is $p = \tilde{x}_u \tilde{a}_d + \tilde{b}_d$.

4.4.1 Per-issue improvement

To inspect the improvement of issue k, for example, we take the sum of the improvement in log-likelihood weighted by each issue:

$$\text{Imp}_k = \frac{\sum_{V_{ud}} \mathbb{E}_q [\boldsymbol{\theta}_{d_v k} | \boldsymbol{w}] (J_{ud} - I_{ud})}{\sum_{V_{ud}} \mathbb{E}_q [\boldsymbol{\theta}_{d_v k} | \boldsymbol{w}]}. \tag{7}$$

A high value of Imp_k indicates that issue k is associated with an increase in log-likelihood, while a low value indicates that the issue saw a decrease in log-likelihood.

Procedural issues such as *Congressional sessions* (in contrast to substantive issues) were among the most-improved issues; they were also much more partisan. This is a result predicted by *procedural cartel theory* [23, 24, 25, 26], which posits that lawmakers will be more polarized in procedural votes (which describe how Congress will be run) than substantive votes (the issues discussed during elections). A substantive issue which was better-predicted was *Finance*, which we illustrate in Figure 4. Infrequent issues like *Women* and *Religion* were nearly unaffected by lawmakers' offsets. In §A.4, we illustrate Imp_k for all issues.

4.4.2 Per-lawmaker improvement

In the 111th House, the per-lawmaker improvement $\text{Imp}_u = \sum_D (J_{ud} - I_{ud})$ was invariably positive or negligible, because each lawmaker has many more parameters in the issue-adjusted model. Some of most-improved lawmakers were Ron Paul and Donald Young.

We corrected lawmakers' issue adjustments to account for their left/right leaning and performed permutation tests as in §4.3 to find which of these corrected adjustments $\hat{z}_{uk}$ were statistically significant at $p < 0.05$ (see supplementary section §A.5 for how we obtain $\hat{z}_{uk}$ from z_{uk} and §A.5 for details on the permutation test). We illustrate these issue adjustments for Paul and Young in Figure 5.

Ron Paul. Paul's offsets were extreme; he voted more conservatively than expected on *Health*, *Human rights* and *International affairs*. He voted more liberally on social issues such as *Racial and ethnic relations*. The issue-adjusted training accuracy of Paul's votes increased from 83.8% to 87.9% with issue offsets, placing him among the two most-improved lawmakers with this model.

The issue-adjusted improvement Imp_K (Equation 7), when restricted to Paul's votes, indicate significant improvement in *International affairs* and *East Asia* (he tends to vote against U.S. involvement in foreign countries); *Congressional sessions*; *Human rights*; and *Special months* (he tends to vote against recognition of special months and holidays). The model hurt performance related to *Law*, *Racial and ethnic relations*, and *Business*, none of which were statistically significant issues for Paul.

Donald Young. One of the most exceptional legislators in the 111th House was Alaska Republican Donald Young. Young stood out in a topic used frequently in House bills about naming local landmarks. Young voted against the majority of his party (and the House in general) on a series of largely symbolic bills and resolutions. In an *Agriculture* topic, Young voted (with only two other Republicans and against the majority of the House) *not* to commend "members of the Agri-business Development Teams of the National Guard [to] increase food production in war-torn countries."

Young's divergent voting was also evident in a series of votes against naming various landmarks–such as post offices–in a topic about such symbolic votes. Notice that Young's ideal point is not particularly distinctive: using the ideal point alone, we would not recognize his unique voting behavior.

4.4.3 Per-bill improvement

Per-bill improvement $\text{Imp}_d = \sum_U (J_{ud} - I_{ud})$ decreased for some bills. The bill which decreased the most from the ideal point model in the 111th House was the *Consolidated Land, Energy, and Aquatic Resources Act of 2010* (H.R. 3534). This bill had substantial weight in five issues, with most in *Public lands and natural resources*, *Energy*, and *Land transfers*, but its placement in many issues harmed our predictions. This effect—worse performance on bills about many issues—suggests that methods which represent bills more sparsely may perform better than the current model.

5 Discussion

Traditional models of roll call data cannot capture how individual lawmakers deviate from their latent position on the political spectrum. In this paper, we developed a model that captures how lawmakers vary, issue by issue, and used the text of the bills to attach specific votes to specific issues. We demonstrated, across 12 years of legislative data, that this model better captures lawmaker behavior. We also illustrated how to use the model as an exploratory tool of legislative data.

Future areas of work include incorporating external behavior by lawmakers. For example, lawmakers make some (but not all) issue positions public. Many raise campaign funds from interest groups. Matching these data to votes would help us to understand what drives lawmakers' positions.

Acknowledgments

We thank the reviewers for their helpful comments. David M. Blei is supported by ONR N00014-11-1-0651, NSF CAREER 0745520, AFOSR FA9550-09-1-0668, the Alfred P. Sloan foundation, and a grant from Google. Yahoo! provided computers for our experiments.

References

[1] Keith T. Poole and Howard Rosenthal. Patterns of congressional voting. *American Journal of Political Science*, 35(1):228–278, February 1991.

[2] Simon Jackman. Multidimensional analysis of roll call data via Bayesian simulation: Identification, estimation, inference, and model checking. *Political Analysis*, 9(3):227–241, 2001.

[3] Joshua Clinton, Simon Jackman, and Douglas Rivers. The statistical analysis of roll call data,. *American Political Science Review*, 98(2):355–370, 2004.

[4] Keith T. Poole and Howard Rosenthal. A spatial model for legislative roll call analysis. *American Journal of Political Science*, pages 357–384, 1985.

[5] Andrew D. Martin and Kevin M. Quinn. Dynamic ideal point estimation via Markov chain Monte Carlo for the U.S. Supreme Court, 1953-1999. *Political Analysis*, 10:134–153, 2002.

[6] Daniel Ramage, David Hall, Ramesh Nallapati, and Christopher D. Manning. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, 2009.

[7] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, pages 993–1022, 2003.

[8] Congressional research service. Available http://www.loc.gov/crsinfo/, 2011.

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

[10] James Albert. Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational Statistics*, 17:251–269, 1992.

[11] James J. Heckman and James M. Snyder. Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators. *RAND Journal of Economics*, 27(0):142–189, 1996.

[12] Matt Thomas, Bo Pang, and Lillian Lee. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the 2006 Conference on Empirical Methods on Natural Language Processing*, 2006.

[13] Tae Yano, Noah A. Smith, and John D. Wilkerson. Textual predictors of bill survival in congressional committees. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics*, page 793802, 2012.

[14] Sean Gerrish and David Blei. Predicting legislative roll calls from text. *Proceedings of the International Conference on Machine Learning*, 2011.

[15] Eric Wang, Dehong Liu, Jorge Silva, David Dunson, and Lawrence Carin. Joint analysis of time-evolving binary matrices and associated documents. *Advances in Neural Information Processing Systems*, 23:2370–2378, 2010.

[16] Chong Wang and David M. Blei. Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th international conference on Knowledge Discovery and Data mining*, pages 448–456, 2011.

[17] Deepak Agarwal and Bee-Chung Chen. fLDA: Matrix factorization through latent Dirichlet allocation. *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, pages 91–100, 2010.

[18] Valen E. Johnson and James H. Albert. *Ordinal Data Modeling*. Springer-Verlag, New York, 1999.

[19] Michael I. Jordan, Zoubin Ghahramani, Tommi S. Jaakkola, and Lawrence K. Saul. An introduction to variational methods for graphical models. *Learning in Graphical Models*, pages 183–233, 1999.

[20] Herbert Robbins and Sutton Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22(3), September 1951.

[21] Leon Bottou and Yann Le Cun. Large scale online learning. In *Advances in Neural Information Processing Systems*, 2004.

[22] Govtrack.us, 2010. Civic Impulse LLC. Available http://www.govtrack.us.

[23] Richard F. Fenno Jr. *The Congress and America's Future*. Prentice-Hall, Englewood Cliffs, NJ, 1965.

[24] Gary W. Cox and Mathew D. McCubbins. *Legislative Leviathon*. University of California Press., 1993.

[25] Gary W. Cox and Keith T. Poole. On measuring partisanship in roll-call voting: The U.S. House of Representatives, 1877-1999. *American Journal of Political Science*, 46(3):pp. 477–489, 2002.

[26] Gary W. Cox and Mathew D. McCubbins. *Setting the Agenda: Responsible Party Government in the U.S. House of Representatives*. Cambridge University Press, 2005.

[27] Harald Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*. Society for Industrial and Applied Mathematics., 1992.

In this appendix we provide additional experiment and implementation details for *How They Vote*.

A.1. Variational posterior inference

We begin by providing more detail about the inference algorithm summarized in the Inference section of the main paper.

Optimizing the variational objective

Variational bounds are typically optimized by gradient ascent or block coordinate ascent, iterating through the variational parameters and updating them until the relative increase in the lower bound is below a specified threshold. Traditionally this would require symbolic expansion of $\mathbb{E}_q\left[p(v, x, \boldsymbol{z}, a, b, \boldsymbol{\theta}) - q(x)\right]$, a process which presupposes familiarity with variational methods.

Instead of expanding this bound symbolically, we update each parameter by making Taylor approximations of the KL objective and performing a series of second-order updates to these parameters, iterating through the parameters and repeating until convergence.

To be concrete, we describe how to perform the mth update on the variational parameter $\tilde{x}$, assuming that we have the most-recent estimate $\tilde{x}_{m-1}$ of this parameter (updates for the other random variables are analogous). Writing the variational objective as $f(\tilde{x}) = \mathrm{KL}(q_{\tilde{x}}||p)$ for notational convenience (where all parameters in η except $\tilde{x}$ are held fixed), we estimate the KL divergence as a function of $\tilde{x}$ around our last estimate $\tilde{x}_{m-1}$ with its Taylor approximation

$$\begin{aligned} f_{m-1}(\tilde{x}) \approx & f(\tilde{x}_{m-1}) + \left(\frac{\partial f}{\partial \tilde{x}}\Big|_{\tilde{x}_{m-1}}\right)^T \Delta\tilde{x} \\ & + \frac{1}{2}\Delta\tilde{x}^T\left(\frac{\partial^2 f}{\partial\tilde{x}\partial\tilde{x}^T}\Big|_{\tilde{x}_{m-1}}\right)\Delta\tilde{x}, \end{aligned} \tag{8}$$

where $\Delta\tilde{x} = \tilde{x} - \tilde{x}_{m-1}$. Once we have estimated the Taylor coefficients (as described in the next section), we can perform the update

$$\tilde{x}_m \leftarrow \tilde{x}_{m-1} - \left(\frac{\partial^2 f_{m-1}}{\partial\tilde{x}\partial\tilde{x}^T}\Big|_{\tilde{x}_{m-1}}\right)^{-1}\left(\frac{\partial f_{m-1}}{\partial\tilde{x}}\Big|_{\tilde{x}_{m-1}}\right). \tag{9}$$

Taylor Coefficient approximation

We approximate the Taylor coefficients in Equation 9 above with Monte Carlo integration, taking samples from $q_{\tilde{x}_{m-1}}$, which is easy to sample from because it has known mean and variance. We approximated the Taylor coefficients by approximating the gradient of $f(\tilde{x}) = \mathrm{KL}(q_{\tilde{x}}||p)$ with samples: We will illustrate this approximate gradient with respect to the variational parameter $\tilde{x}$. Let $\tilde{x}_0$ be the current estimates of the variational mean, $q_{\tilde{x}_0}(x, \boldsymbol{z}, a, b)$ be the variational posterior at this mean, and define $\mathcal{L}_{\tilde{x}_0} = \mathbb{E}_q\left[p(x_0, \boldsymbol{z}, a, b) - q(x_0, \boldsymbol{z}, a, b)\right]$.

We then approximate the gradient with Monte Carlo samples as

$$\frac{\partial \mathcal{L}_{\tilde{x}_0}}{\partial \tilde{x}}\Big|_{\tilde{x}_0} = \frac{\partial}{\partial \tilde{x}} \int q_{\tilde{x}}(x, \boldsymbol{z}, a, b)(\log p(x, \boldsymbol{z}, a, b, v) \tag{10}$$

$$- \log q_{\tilde{x}}(x, \boldsymbol{z}, a, b))dxd\boldsymbol{z}dadb \tag{11}$$

$$= \int \frac{\partial}{\partial \tilde{x}} \left(q_{\tilde{x}}(x)(\log p(x, \boldsymbol{z}, a, b, v) - \log q_{\tilde{x}}(x, \boldsymbol{z}, a, b))\right) d\tilde{x}$$

$$= \int q_{\tilde{x}}(x) \frac{\partial \log q_{\tilde{x}(x)}}{\partial \tilde{x}} (\log p(x, \boldsymbol{z}, a, b, v) - \log q_{\tilde{x}}(x, \boldsymbol{z}, a, b))d\tilde{x}$$

$$\approx \frac{1}{N}\Bigg(\sum_{n=1}^{N} \frac{\partial \log q_{\tilde{x}}(x_n, z_n, a_n, b_n)}{\partial \tilde{x}}\Big|_{\tilde{x}_0} \times \big(\log p(x_n, z_n, a_n b_n, v) - C - \log q_{\tilde{x}_0}(x_n, z_n, a_n, b_n)\big)\Bigg),$$

where we have used N samples from the current estimate of the variational posterior.

$$\frac{\partial f}{\partial \tilde{x}}\Big|_{\tilde{x}_{m-1}} = \frac{\partial}{\partial \tilde{x}} \int q_{\tilde{x}}(x)(\log p(x) - \log q_{\tilde{x}}(x))d\tilde{x} \tag{12}$$

$$= \int \frac{\partial}{\partial \tilde{x}} \left(q_{\tilde{x}}(x)(\log p(x) - \log q_{\tilde{x}}(x))\right) d\tilde{x}$$

$$= \int q_{\tilde{x}}(x) \frac{\partial \log q_{\tilde{x}(x)}}{\partial \tilde{x}} (\log p(x) - \log q_{\tilde{x}}(x))d\tilde{x}$$

$$\approx \frac{1}{N}\Bigg(\sum_{n=1}^{N} \frac{\partial \log q_{m-1}(x_{m-1,n})}{\partial \tilde{x}_{m-1}}\Big|_{\tilde{x}_{m-1}} \times \big(\log p(x_{m-1,n}|\boldsymbol{z}_{m-1,n}, x_{m-1,n}, a_{m-1,n}, b_{m-1,n}, V) - C - \log q_{m-1}(x_{m-1,n})\big)\Bigg),$$

where we have taken the gradient through the integral using Liebniz's rule. The second Taylor coefficient is straightforward to derive with similar algebra:

$$\begin{aligned}\frac{\partial^2 f}{\partial \tilde{x} \partial \tilde{x}^T}\Big|_{\tilde{x}_{m-1}} \approx \frac{1}{N} \sum_{n=1}^{N} \Bigg(&\left(\frac{\partial \log q_{m-1}(x_{m-1,n})}{\partial \tilde{x}}\Big|_{\tilde{x}_{m-1}} \right) \times \left(\frac{\partial \log q_{m-1}(x_{m-1,n})}{\partial \tilde{x}}\Big|_{\tilde{x}_{m-1}} \right)^T \\ &\times \big(\log p(x_{m-1,n}|\boldsymbol{z}_{m-1,n}, a_{m-1,n}, b_{m-1,n}, V) - C - \log q_{m-1}(x_{m-1,n}) - 1\big) \\ &+ \Bigg(\left(\frac{\partial^2 \log q_{m-1}(x_{m-1,n})}{\partial \tilde{x} \partial \tilde{x}^T}\Big|_{\tilde{x}_{m-1}} \right) \times \big(\log p(x_{m-1,n}|\boldsymbol{z}_{m-1,n}, a_{m-1,n}, b_{m-1,n}, V) - C - \log q_{m-1}(x_{m-1,n})\big)\Bigg)\Bigg),\end{aligned} \tag{13}$$

where we increase N as the model converges. Note that C is a free parameter that we can set without changing the final solution. We set C to the average of $\log p(x_{m-1,n}|...) - \log q_{m-1}(x_{m-1,n})$ across the set of N samples.

Instead of taking *iid* samples from the variational distribution q_{M-1}, we used quasi-Monte Carlo sampling [27]. By taking non-*iid* samples from q_{m-1}, we are able to decrease the variance around estimates of the Taylor coefficients. To select these samples, we took N equally-spaced points from the unit interval, passed these through the inverse CDF of the variational Gaussian $q_{m-1}(x)$, and used the resulting values as samples.[6]

We did this for each random variable in the Markov blanket of x_u, permuted each variable's samples, and combined them for N multivariate samples $\{x_{m-1,n}, \ldots, B_{m-1,n}\}_n$ from the current estimate q_{m-1} of the variational distribution.

We estimate the gradients of $\log q$ above based on the distribution of the variational marginals. We have defined the variational distribution to be factorized Gaussians, so these take the form

$$\begin{aligned} \frac{\partial \log q_{m-1}(x_{m-1,n})}{\partial \tilde{x}}\Big|_{\tilde{x}_{m-1}} &= \frac{x_{m-1,n} - \tilde{x}_{m-1}}{\sigma_x^2} \\ \frac{\partial^2 \log q_{m-1}(x_{m-1,n})}{\partial \tilde{x}^2}\Big|_{\tilde{x}_{m-1}} &= -\frac{1}{\sigma_x^2}. \end{aligned} \tag{14}$$

The variance σ_x^2 was fixed to $\exp(-5)$. Allowing σ_x to vary freely provides a better variational bound at the expense of accuracy. This happens because the issue-adjusting model would sometimes fit poor means to some parameters when the posterior variance was large: there is little penalty for this when the variance is large. Low posterior variance σ_x^2 is similar to a non-sparse MAP estimate.

These updates were repeated until the exponential moving average $\Delta_{\text{est},i} \leftarrow 0.8\Delta_{\text{est},i-1} + 0.2\Delta_{\text{obs},i}$ of the change in KL divergence dropped below one and the number N of samples passed 500. If the moving average dropped below one and $N < 500$, we doubled the number of samples.

For all experiments, we began with $M = 21$ samples to estimate the approximate gradient and scaled it by 1.2 each time the Elbo dropped below a threshold, until it passed 500.

Sparsity.

Issue adjustments $\boldsymbol{z}_u$ ranged widely, moving some lawmakers significantly. The variational estimates were not sparse, although a high mass was concentrated around 0. Twenty-nine percent of issue adjustments were within $[-0.01, 0.01]$, and eighty-seven percent of issue adjustments were within $[-0.1, 0.1]$.

Numerical stability and hyperparameter sensitivity

We address practical details of implementing issue-adjusted ideal points.

Hyperparameter settings

The most obvious parameter in the issue voting model is the regularization term λ. The Bayesian treatment described in the Inference section of *How they Vote* demonstrated considerable robustness to overfitting at the expense of precision. With $\lambda = 0.001$, for example, issue adjustments z_{uk} remained on the order of single digits, while the MAP estimate yielded adjustment estimates over 100.

We recommend a modest value of $1 < \lambda < 10$. At this value, the model outperforms ideal points in validation experiments consistently in both the House and Senate.

Implementation.

When performing the second-order updates described in the Inference section, we skipped variable updates when the estimated Hessian was not positive definite (this disappeared when sample sizes grew large enough). We also limited step sizes to 0.1 (another possible reason for smaller coefficients).

[6]Note that these samples produce a biased estimate of Equation 8. This bias decreases as $N \to \infty$.

A.2. Issue labels

In the empirical analysis, we used issue labels obtained from the Congressional Research Service. There were $5,861$ labels, ranging from *World Wide Web* to *Age*. We only used issue labels which were applied to at least twenty five bills in the 12 years under consideration. This filter resulted in seventy-four labels which correspond fairly well to political issues. These issues, and the number of documents each label was applied to, is given in Table 2.

Table 2: Issue labels and the number of documents with each label (as assigned by the Congressional Research Service) for Congresses 106 to 111 (1999 to 2010).

Issue label	**Bills**
Women	25
Military history	25
Civil rights	25
Government buildings; facilities; and property	26
Terrorism	26
Energy	26
Crime and law enforcement	27
Congressional sessions	27
East Asia	28
Appropriations	28
Business	29
Congressional reporting requirements	30
Congressional oversight	30
Special weeks	31
Social services	31
Health	33
Special days	33
California	33
Social work; volunteer service; charitable organizations	33
State and local government	34
Civil liberties	35
Government information and archives	35
Presidents	35
Government employees	35
Executive departments	35
Racial and ethnic relations	36
Sports and recreation	36
Labor	36
Special months	39
Children	40
Veterans	40
Human rights	41
Finance	41
Religion	42
Politics and government	43
Minorities	44
Public lands and natural resources	44

Issue label	**Bills**
Europe	44
Military personnel and dependents	44
Taxation	47
Government operations and politics	47
Postal facilities	47
Medicine	48
Transportation	48
Emergency management	48
Sports	52
Families	53
Medical care	54
Athletes	56
Land transfers	56
Armed forces and national security	56
Natural resources	58
Law	60
History	61
Names	62
Criminal justice	62
Communications	65
Public lands	68
Legislative rules and procedure	69
Elementary and secondary education	74
Anniversaries	82
Armed forces	83
Defense policy	92
Higher education	103
Foreign policy	104
International affairs	105
Budgets	112
Education	122
House of Representatives	142
Commemorative events and holidays	195
House rules and procedure	329
Commemorations	400
Congressional tributes	541
Congress	693

Corpus preparation

In this section we provide further details of vocabulary selection. We began by considering all phrases with one to five words. From these, we immediately ignored phrases which occurred in more than 10% of bills and fewer than 4 bills, or which occurred as fewer than 0.001% of all phrases. This resulted in a list of 40603 phrases.

Table 3: Features and coefficients used for predicting “good” phrases. Below, test is a test statistic which measures deviation from a model assuming that words appear independently; large values indicate that they occur more often than expected by chance. We define it as $\text{test} = \frac{\text{Observed count} - \text{Expected count}}{\sqrt{\text{Expected count under a language model assuming independence}}}$

Coefficient	Summary	Weight
log(count + 1)	Frequency of phrase in corpus	-0.018
log(number.docs + 1)	Number of bills containing phrase	0.793
anchortext.presentTRUE	Occurs as anchortext in Wikipedia	1.730
anchortext	Frequency of appearing as anchortext in Wikipedia	1.752
frequency.sum.div.number.docs	Frequency divided by number of bills	-0.007
doc.sq	Number of bills containing phrase, squared	-0.294
has.secTRUE	Contains the phrase *sec*	-0.469
has.parTRUE	Contains the phrase *paragra*	-0.375
has.strikTRUE	Contains the phrase *strik*	-0.937
has.amendTRUE	Contains the phrase *amend*	-0.484
has.insTRUE	Contains the phrase *insert*	-0.727
has.clauseTRUE	Contains the phrase *clause*	-0.268
has.provisionTRUE	Contains the phrase *provision*	-0.432
has.titleTRUE	Contains the phrase *title*	-0.841
test.pos	$\ln(max(-\text{test}, 0) + 1)$	0.091
test.zeroTRUE	1 if test $= 0$	-1.623
test.neg	$\ln(max(\text{test}, 0) + 1)$	0.060
number.terms1	Number of terms in phrase is 1	-1.623
number.terms2	Number of terms in phrase is 2	2.241
number.terms3	Number of terms in phrase is 3	0.315
number.terms4	Number of terms in phrase is 4	-0.478
number.terms5	Number of terms in phrase is 5	-0.454
log(number.docs + 1) * anchortext	$\ln(\text{Number of bills containing phrase}) \times 1_{\{\text{Appears in Wikipedia anchortext}\}}$	-0.118
log(count + 1) * log(number.docs + 1)	$\ln(\text{Number of bills containing phrase} + 1) \times \ln(\text{Frequency of phrase in corpus} + 1)$	0.246

We then used a set of features characterizing each word to classify whether it was good or bad to use in the vocabulary. Some of these features were based on corpus statistics, such as the number of bills in which a word appeared. Other features used external data sources, including whether, and how frequently, a word appeared as link text in a Wikipedia article. For training data, we used a manually curated list of 458 “bad” phrases which were semantically awkward or meaningles (such as *the follow bill*, *and sec ammend*, *to a study*, and *pr*) as negative examples in a L_2-penalized logistic regression to select a list of 5,000 “good” words.

A.3. Summary of corpus statistics

We studied U.S. Senate and House of Representative roll-call votes from 1999 to 2010. This period spanned Congresses 106 to 111 and covered an historic period in U.S. history, the majority of which Republican President George W. Bush held office. Bush’s inauguration and the attacks of September 11th, 2001 marked the first quarter of this period, followed by the wars in Iraq and Afghanistan. Democrats gained a significant share of seats from 2007 to 2011, taking the majority from Republicans in both the House and the Senate, and Democratic President Obama was inaugurated in January 2009. A summary of statistics for our datasets in these Congresses is provided in Table 4.

Table 4: Roll-call data sets used in the experiments. These counts include votes in both the House and Senate. Congress 107 had fewer fewer votes than the remaining congresses in part because this period included large shifts in party power, in addition to the attacks on September 11th, 2001.

Congress	Years	Lawmakers	Bills	Votes (Senate)
106	1999-2000	516	391	149,035 (7,612)
107	2001-2002	391	137	23,996 (5,547)
108	2003-2004	539	527	207,984 (7,830)
109	2005-2006	540	487	194,138 (7,071)
110	2007-2008	549	745	296,664 (9,019)
111	2009-2010	552	826	336,892 (5,936)

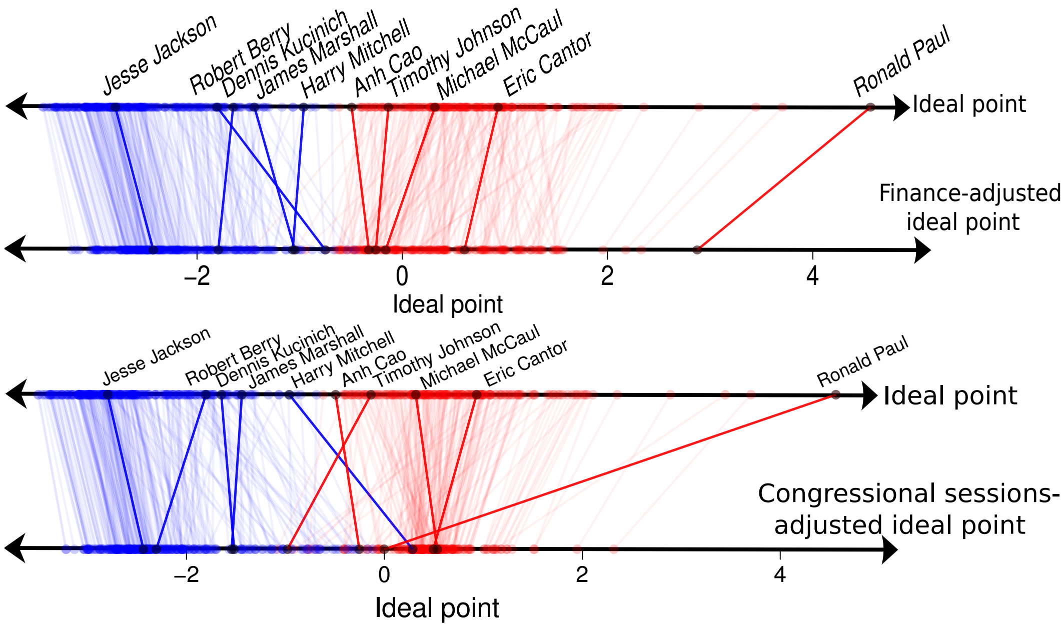


Figure 6: Ideal points x_u and issue-adjusted ideal points $x_u + z_{uk}$ from the 111th House. Democrats are blue and Republicans are red. Votes were most improved for the issue *Congressional sessions*, which focuses on procedural matters such as when to adjourn for a House recess or whether to consider certain legislation. Lawmakers were split into factions: some became further polarized by these bills, but some did not; the resulting mixture was not on party lines. Votes about Finance were also better fit with this model. Democrats were mostly fixed on this issue, but Republicans (who were less-well predicted by ideal points alone) saw more adjustment.

A.4. Additional figures

Figure 6 shows lawmakers offsets for two different issues. This exemplifies how much lawmakers diverge from a one-dimensional ideal point model.

Figure 7 illustrates the extent to which the issue-adjusted ideal point model improves prediction for different issues. These values were computed over a fit of the model to all votes in the 111th House of Representatives.

A.5. Controlling for lawmakers' ideal point x_u in issue adjustments

Controlling for ideal points

The issue-adjusted ideal point model is under-specified in several ways. It is well known that the signs of ideal points x_u and bill polarities a_d are arbitrary, for example, because $x_u a_d = (-x_u)(-a_d)$.

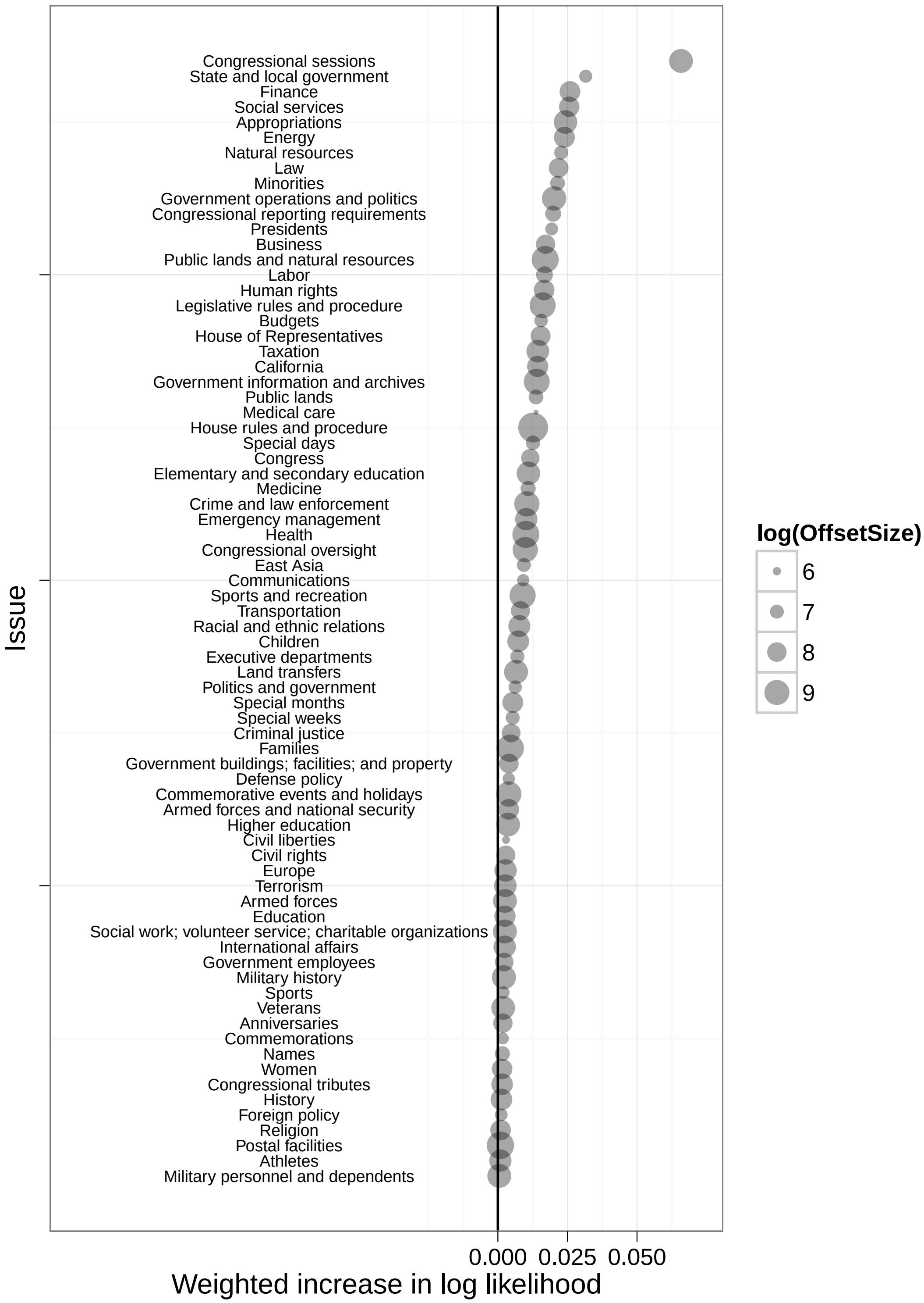


Figure 7: Issue adjustments (defined in Equation 6) for all issues.

This leads to a multimodal posterior [2]. We address this by flipping ideal points (and bill polarities) if necessary to make Republicans positive and Democrats negative.

The model is also underspecified because lawmakers' issue preferences can be explained in part by their ideal points (this is especially true on procedural issues). A typical Republican tends to have a Republican offset on taxation, but this surprises nobody. Instead, we are more interested in understanding when this Republican lawmaker *deviates* from behavior suggested by his ideal point. We therefore fit a regression for each issue k to explain away the effect of a lawmaker's ideal point $\boldsymbol{x}_u$ on her offset $\boldsymbol{z}_{uk}$:

$$\boldsymbol{z}_k = \beta_k \boldsymbol{x} + \boldsymbol{\varepsilon},$$

where $\beta_k \in \mathbb{R}$. Instead of evaluating a lawmaker's observed offsets, we use her residual $\hat{z}_{uk} = \boldsymbol{z}_{uk} - \beta_k \boldsymbol{x}_u$. By doing this, we can evaluate lawmakers in the context of other lawmakers who share the same ideal points: a positive offset $\hat{z}_{uk}$ for a Democrat means she tends to vote more liberally about issue k than Democrats with the same ideal point.[7]

Most issues had only a moderate relationship to ideal points. *House rules and procedure* was the most-correlated with ideal points, moving the adjusted ideal point $\beta_k = 0.26$ right for every unit increase in ideal point. *Public land and natural resources* and *Taxation* followed at a distance, moving an ideal point 0.04 and 0.025 respectively with each unit increase in ideal point. *Health*, on the other hand, moved lawmakers $\beta_k = 0.04$ left for every unit increase in ideal point.

Assessing significance

A handful of lawmakers stood out with the most exceptional issue adjustments. Any reference in this section to lawmakers' issue adjustments refers to lawmakers' residuals $\hat{z}_{uk}$ fit from their variational parameters $\tilde{z}_{uk}$. Lawmakers' issue adjustments are confounded because estimated issue adjustments had high variance, and issue adjustments had fatter tails than expected under a normal distribution. We therefore turned again to the same nonparametric permutation test described in the main experiments section: permute issue vectors' document labels, i.e. $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_D) \mapsto (\boldsymbol{\theta}_{\pi_i(1)} \ldots \boldsymbol{\theta}_{\pi_i(D)})$, and refit lawmakers' adjustments using both the original issue vectors and permuted issue vectors. We then compare a normal issue residual $\hat{z}_{uk}$'s absolute value with issue residuals $\hat{z}_{uki}$ estimated with permuted issue vectors $\boldsymbol{\theta}_{\pi_i(d)k}$. By performing this test twenty times, we can say that a lawmaker's offset $\hat{z}_{uk}$ is significant if it is outside of the range of $\{\hat{z}_{uki}\}_i$ for all permutations i.

This provides a nonparametric method for finding issue adjustments which are more extreme than expected by chance: an extreme issue adjustment has a greater absolute value than all of its permuted counterparts.

[7] We also fit a model with this regression explicitly encoded. That model performed slightly worse in experiments on heldout data.