The Utility of Text: The Case of Amicus Briefs and the Supreme Court

Yanchuan Sim
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA 15213
yanchuan@cmu.edu

Bryan Routledge
Tepper School of Business
Carnegie Mellon University
Pittsburgh, PA 15213
routledge@cmu.edu

Noah A. Smith
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA 15213
nasmith@cs.cmu.edu

Abstract

We explore the idea that authoring a piece of text is an act of maximizing one's expected utility. To make this idea concrete, we consider the societally important decisions of the Supreme Court of the United States. Extensive past work in quantitative political science provides a framework for empirically modeling the decisions of justices and how they relate to text. We incorporate into such a model texts authored by amici curiae ("friends of the court" separate from the litigants) who seek to weigh in on the decision, then explicitly model their goals in a random utility model. We demonstrate the benefits of this approach in improved vote prediction and the ability to perform counterfactual analysis.

1 Introduction

Some pieces of text are written with clear goals in mind. Economists and game theorists use the word *utility* for the concept of satisfying a need or desire, and a huge array of theories and models are available for analyzing how utility-seeking agents behave. This paper takes the first steps in incorporating *text* into these models.

The Supreme Court of the United States (SCOTUS) is the highest court in the American judicial system; its decisions have far-reaching effects. While the ideological tendencies of justices (SCOTUS' nine justices) are widely discussed by press and public, there is a formal mechanism by which organized interest groups can lobby the court on a given case. These groups are known as *amici curiae* (Latin for "friends of the court," hereafter "amici" and singular "amicus"), and the textual artifacts they author—known as amicus briefs—reveal explicit attempts to sway justices one way or the other. Taken alongside voting records and other textual artifacts that characterize a case, amicus briefs provide a fascinating setting for the empirical study of influence through language.

We build on a well-established methodology from political science known as *ideal points* for analyzing votes. Specifically, Lauderdale and Clark (2014) combined descriptive text and ideal points in a probabilistic topic model. Although the influence of amici has been studied extensively by legal scholars (Collins, 2008), we are the first to incorporate them into ideal points analysis (§2). Drawing on decision theory, we then posit amici as rational agents seeking to maximize their expected utility by framing an amicus brief's arguments to influence justices toward a favorable outcome (§3). We derive appropriate inference and parameter estimation procedures (§4).

Our experiments (§6) show that the new approach offers substantial gains in vote prediction accuracy. More importantly, we show how the model can be used to answer questions such as: How susceptible are different justices to influence by amici? How effective were amici on each side of a case? What would have happened if some or all amicus briefs were not filed? How might an amicus have changed her brief to obtain a better outcome? Since our approach characterizes the amicus brief, as a (probabilistic) function of the case parameters, our approach could also be used to ask how the amici would have altered their briefs given different merits facts or a different panel of justices. Although we focus on SCOTUS, our model is applicable to any setting where textual evidence for competing goals are available alongside behavioral response.

SCOTUS Terminology. SCOTUS reviews the decisions of lower courts and (less commonly) resolves disputes between states. The primary means to petition SCOTUS to review a case is to ask it to grant a writ of certiorari; this happens in about 1.5% of 7,000 petitions a year. SCOTUS is under no obligation to hear a case.[1] Once accepted, the **petitioner** writes a brief putting forward her legal argument; the **respondent** (the other party) then files a brief. These, together with a round of responses to each other's initial briefs, are collectively known as **merits briefs**. **Amicus briefs**—further arguments and recommendations on either side—may be filed by groups with an interest (but not a direct stake) in the outcome, with the Court's permission. After oral arguments (which are not allotted for every case) conclude, the justices vote and author one or more opinions. A justice from the majority is assigned to author the rationale for the court's decision; others may write separate concurring or dissenting opinions. In this paper, we will be studying how the votes of justices relate to merits and amicus briefs.

2 Ideal Point Models

Ideal point (IP) models are a mainstay in quantitative political science, often being applied to voting records to place voters (lawmakers, justices, etc.) in a continuous space. A justice's "ideal point" is a latent variable positioning her in this space.

2.1 Unidimensional Ideal Points

The simplest model for judicial votes is a unidimensional IP model (Martin and Quinn, 2002), which posits an IP $\psi_j \in \mathbb{R}$ for each justice j.[2] Often the ψ_j values are interpreted as positions along a liberal-conservative ideological spectrum. Each case i is represented by *popularity* (a_i) and *polarity* (b_i) parameters.[3] A probabilistic view of the unidimensional IP model is that justice j votes in favor of case i's petitioner with probability

$$p(v_{i,j} = \text{petitioner} \mid \psi_j, a_i, b_i) = \sigma\left(a_i + \psi_j b_i\right)$$

where $\sigma(x) = \frac{\exp(x)}{1+\exp(x)}$ is the logistic function. When the popularity parameter a_i is high enough, every justice is more likely to favor the petitioner. The polarity b_i captures the importance of the justice's ideology

[1]Details about the procedures and rules of the SCOTUS can be found at `http://www.uscourts.gov`, according to which a case is heard "if the case could have national significance, might harmonize conflicting decisions in the federal Circuit courts, and/or could have precedential value."

[2]Martin and Quinn (2002) describe a dynamic unidimensional IP model where justice IP vary over time. In this work, we assume each justice's IP is fixed over time, for simplicity.

[3]This model is also known as a two parameter logistic model in item response theory literature (Fox, 2010), where a_i is "difficulty" and b_i is "discrimination."

ψ_j: more polarizing cases (i.e., $|b_i| \gg 0$) push justice j more strongly to the side of the petitioner (if b_i has the same sign as ψ_j) or the respondent (otherwise).

The unidimensional IP model has been extended into multiple dimensions using the same formulation.[4] While IP models recover dimensions that maximize statistical fit, they conflate many substantive dimensions of opinion and policy, making it difficult to interpret additional dimensions.[5] Indeed, such embeddings are ignorant of the issues at stake, or any content of the case, and they cannot generalize to new cases.

2.2 Issues and Ideal Points

Lauderdale and Clark (2014) incorporate text as evidence and infer dimensions of IP that are grounded in "topical" space. They build on latent Dirichlet allocation (Blei et al., 2003), a popular model of latent topics or themes in text corpora. In their model, each case i is embedded as $\boldsymbol{\theta}_i$ in a D-dimensional simplex; the dth dimension $\theta_{i,d}$ corresponds to the proportion of case i that is about issue (or, in LDA terminology, topic) d. The probability of justice j's vote is given by

$$p(v_{i,j} = \text{pet'r} \mid \boldsymbol{\psi}_j, \boldsymbol{\theta}_i, a_i, b_i) = \sigma\left(a_i + \boldsymbol{\psi}_j^\top \left(b_i \boldsymbol{\theta}_i\right)\right)$$

where $\psi_{j,d}$ is an *issue-specific* position for justice j. Therefore, the relative degree that each dimension predicts the vote outcome is determined by the text's mixture proportions, resulting in the issue-specific IP $\boldsymbol{\psi}_j^\top \boldsymbol{\theta}_i$. In their work, they inferred the mixture proportions from justices' opinions, although one can similarly use merits briefs, appeals court opinions, or any other texts that serve as evidence for inferring the issues of a case.

Lauderdale and Clark (2014) found that incorporating textual data in this manner[6] addresses the labeling problem for multidimensional models, and is especially useful for small voting bodies (i.e. SCOTUS), where estimating multidimensional models is difficult due to few observations and variation of preferences across issues.

2.3 Amici and Ideal Points

The merits briefs describe the issues and facts of the case. It is our hypothesis that amicus briefs serve to "frame" the facts and, potentially, influence the case outcome. Collins (2008) argued that these organized interest groups play a significant role in shaping justices' choices. Likewise, Corley et al. (2013) have observed that justices systematically incorporate language from amicus briefs to enhance their ability to make effective law and policy. For instance, SCOTUS often requests the Solicitor General to file amicus briefs, attesting to the briefs' importance. Public interest groups, such as the American Civil Liberties Union, and Citizens United, frequently advocate their positions on any cases that impinges on their goals. These briefs can provide valuable assistance to the Court in its deliberation; for example, they can present an argument not found in the merits.[7]

When filing amicus briefs, amici are required to identify the side they are supporting, or indicate whether they suggest affirmance or reversal, or if they support neither side. However, it may not always be trivial

[4]See Jackman (2001) and Martin and Quinn (2001) for examples of traditional multidimensional IP; also §7.

[5]A seminal finding of Poole and Rosenthal (1985) is that two dimensions, corresponding to left-right ideology and geographical latitude, explain most of the variance in U.S. Congressional votes.

[6]Of course, LDA is not the only way to "embed" a case in a simplex. One can take advantage of expert categorization of case issues. For example, Gerrish and Blei (2012) used bill labels as supervision to infer the proportions of issues.

[7]On occasion, SCOTUS may adopt a position not advanced by either side, but instead urged solely by an amicus brief. Some notable cases are: *Mapp v. Ohio*, 367 U.S. 643, 646 (1961), regarding exclusionary rule in cases of Fourth Amendment violations, raised exclusively by amicus ACLU; more recently, in *Turner v. Rogers*, 131 S. Ct. 2507 (2011), the Court based its decision on an argument urged only in an amicus brief submitted by the Solicitor General.

to tell which side the amici are on as these intentions are found in the brief text and are not expressed consistently. We solve this by training a classifier on hand-labeled data (§5).

We propose that amici represent an attempt to shift the position of the case by emphasizing some issues more strongly or framing it distinctly from the perspectives given in the merits briefs. The effective position of a case, previously $b_i\boldsymbol{\theta}_i$, is in our model $b_i\boldsymbol{\theta}_i + c_i^{\mathrm{p}}\boldsymbol{\Delta}_i^{\mathrm{p}} + c_i^{\mathrm{r}}\boldsymbol{\Delta}_i^{\mathrm{r}}$, where c_i^{p} and c_i^{r} are the *amicus polarities* for briefs on the side of the petitioner and respondent. $\boldsymbol{\Delta}_i^{\mathrm{p}}$ and $\boldsymbol{\Delta}_i^{\mathrm{r}}$ are the mean issue proportions of the amicus briefs on the side of the petitioner and respondent, respectively. Our amici-augmented IP model is:

$$\begin{aligned} &p(v_{i,j} = \text{petitioner} \mid \psi_j, \boldsymbol{\theta}_i, \boldsymbol{\Delta}_i, a_i, b_i, c_i) \\ &= \sigma\left(a_i + \boldsymbol{\psi}_j^\top \left(b_i\boldsymbol{\theta}_i + c_i^{\mathrm{p}}\boldsymbol{\Delta}_i^{\mathrm{p}} + c_i^{\mathrm{r}}\boldsymbol{\Delta}_i^{\mathrm{r}}\right)\right) \end{aligned} \tag{1}$$

In this model, the vote-specific IP is influenced by two forms of text: legal arguments put forth by the parties involved (merits briefs, embedded in $\boldsymbol{\theta}_i$), and by the amici curiae (amicus briefs, embedded in $\boldsymbol{\Delta}_i^{\{\mathrm{p,r}\}}$), both of which are rescaled independently by the case discrimination parameters to generate the vote probability. When either $|c_i^{\mathrm{p}}|$ or $|c_i^{\mathrm{r}}|$ is large (relative to a_i and b_i), the vote is determined by the contents of the amicus briefs. Hereafter, we let $\boldsymbol{\kappa}_i = \langle a_i, b_i, c_i^{\mathrm{p}}, c_i^{\mathrm{r}}\rangle$.

By letting $\boldsymbol{\Delta}_i^s$ be the average mixture proportions inferred from text of briefs supporting the side s, we are implicitly assuming that briefs supporting the same side share a single parameter, and individual briefs on one side influence the vote-specific IP equally. While Lynch (2004) and others have argued that some amici are more effective (i.e., influence on justices' votes varies across amicus authors), our model captures the collective effect of amicus briefs and is simple.

3 Amici as Agents

In the previous section, the IP models focus on justices' positions embedded in a continuous space. However, we want to account for the fact that amici are purposeful decision makers who write briefs hoping to sway votes on a case. Suppose we have an amicus curiae supporting side s (e.g., petitioner), which is presided by a set of justices, $\mathcal{J}$. The amicus is interested in getting votes in favor of her side, that is, v_j=s. Thus, we assume that she has a simple evaluation function over the outcome of votes $v_1, \ldots, v_9$, i.e., her **utility** is

$$u(v_1, v_2, \ldots, v_9) = \textstyle\sum_{j\in\mathcal{J}} \mathbb{I}(v_j = s) \tag{2}$$

where $\mathbb{I}$ is the indicator function.

Cost of writing. In addition to the policy objectives of an amicus, we need to characterize her "technology" (or "budget") set. We do this here by specifying a cost function, C, that is increasing in difference between $\boldsymbol{\Delta}$ and the "facts" in $\boldsymbol{\theta}$:

$$C(\boldsymbol{\Delta}, \boldsymbol{\theta}) = \tfrac{\xi}{2}\|\boldsymbol{\Delta} - \boldsymbol{\theta}\|_2^2$$

where $\xi > 0$ is a hyperparameter controlling the cost (relative to the vote evaluation). The function captures the notion that amicus briefs cannot be arbitrary text; there is disutility or effort required to carefully frame a case, or the monetary cost of hiring legal counsel. The key assumption here is that framing is costly, while simply matching the merits is easy (and presumably unnecessary). Note the role of the cost function is analogous to regularization in other contexts.

Expected utility. Since the outcome of the case is uncertain, the amicus' objective will consider her *expected* utility:[8]

$$\max_{\boldsymbol{\Delta}} \ \mathbb{E}_{\boldsymbol{\Delta}}[u(v_1,\ldots,v_9)] - \tfrac{\xi}{2}\|\boldsymbol{\Delta}-\boldsymbol{\theta}\|_2^2$$

When an amicus writes her brief, we assume that she has knowledge of the justices' IPs, case parameters, and contents of the merits briefs, but ignores other amici.[9] As such, taking linearity of expectations, we can compute the expected utility for an amicus on side s using Eq. (1):

$$\max_{\boldsymbol{\Delta}} \ \textstyle\sum_{j\in\mathcal{J}} \sigma\left(a + \boldsymbol{\psi}_j^\top(b\boldsymbol{\theta} + c^s\boldsymbol{\Delta})\right) - \tfrac{\xi}{2}\|\boldsymbol{\Delta}-\boldsymbol{\theta}\|_2^2$$

Brief writing trade-offs. Consider the first-order (or KKT) condition for the purposeful amicus' maximization w.r.t. the dth issue:[10]

$$\begin{aligned}&\textstyle\sum_{j\in\mathcal{J}} \sigma'\psi_{j,d}c^s - \xi(\Delta_d-\theta_d)\\ &= \textstyle\sum_{j\in\mathcal{J}} \sigma'\psi_{j,1}c^s - \xi(\Delta_1-\theta_1) \qquad\qquad \text{if } 0<\Delta_d<1\end{aligned}$$

where σ' is the first order derivative of the vote probability sigmoid function. Optimality drives all topics (with positive weight) to have equal marginal values.[11] The marginal value highlights the tradeoffs an amicus faces, in four components: (i) the cost of deviating from the merits, i.e., a large difference between Δ_d and θ_d, is ξ; (ii) a justice whose σ' is large, i.e., whose vote is uncertain, receives greater attention, in particular (iii) on issues she cares about, i.e., $\psi_{j,d}$ is large, with (iv) c^s controlling the strength of correspondence to issues justices care about.

Random utility models. There are several conceivable ways to incorporate amici's optimization into our estimation of justices' IP. We could maximize the likelihood and impose the constraint on $\boldsymbol{\Delta}$ that solve our expected utility optimization (either directly or by checking the first order conditions).[12] Or, we can view such (soft) constraints as imposing a prior on $\boldsymbol{\Delta}$:

$$p_{\text{util}}(\boldsymbol{\Delta}) \propto \mathbb{E}_{\boldsymbol{\Delta}}[u(v_1,\ldots)] + \xi(1-\tfrac{1}{2}\|\boldsymbol{\Delta}-\boldsymbol{\theta}\|_2^2) \tag{3}$$

where the constant is added so that p_{util} is non-negative. Note, were the utility negative, the amici would have chosen not to write a the brief. This approach is known as a **random utility model** in the econometrics discrete-choice literature (McFadden, 1974). Random utility models relax the precision of the optimization by assuming that agent preferences also contain an idiosyncratic random component. Hence, the behavior we observe, (i.e. the amicus' topic mixture proportions), has a likelihood that is proportional to expected utility. Considering all the amici, the full likelihood we estimate is

$$\mathcal{L}(\boldsymbol{w},\boldsymbol{v},\boldsymbol{\psi},\boldsymbol{\theta},\boldsymbol{\Delta},\boldsymbol{\kappa}) \times \left[\textstyle\prod_{k\in\mathcal{A}} p_{\text{util}}(\boldsymbol{\Delta}_k)\right]^\eta \tag{4}$$

where $\mathcal{L}(\cdot)$ is the likelihood of our amici IP model (Eq. 1), and η is a hyperparameter that controls influence of utility on parameter estimation.

---

[8] We use an evaluation function that is linear in votes for simplicity. The scale of the function is unimportant (expected utility is invariant to affine transformations). However, we leave for future work other specifications of the evaluation function; for example a function that places more emphasis on the majority vote outcome.

[9] A model with strategic amici agents (a petitioner amicus choosing brief topics considering a respondent amicus' brief) is a very complicated game theoretical model and, we conjecture, would reequire a much richer representation of policy and goals.

[10] We consider inner points for clarity; if $\Delta_d = 1$, equality is replaced with $\geq$ (respectively, 0 and $\leq$).

[11] Comparing each topic's proportion to topic 1 is arbitrary. In a D-topic model, the amicus has only $D-1$ choices.

[12] This is reminiscent of learning frameworks where constraints are placed on the posterior distributions (Chang et al., 2007; Ganchev et al., 2010; McCallum et al., 2007). However, the nonlinear nature of our expectations makes it difficult to optimize and characterize the constrained distribution.

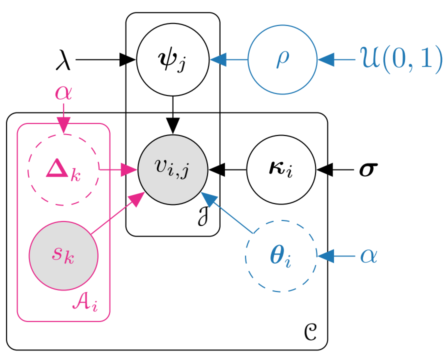
Eq. (4) resembles *product of experts* model (Hinton, 2002). For the likelihood of votes in a case to be maximized, it is necessary that no individual component—generative story for votes, amicus briefs—assigns a low probability. Accordingly, this results in a principled manner for us to incorporate our assumptions about amici as rational decision makers, each of whom is an "expert" with the goal of nudging latent variables to maximize her own expected utility.

4 Learning and Inference

Model priors. The models we described above combine IP, topic models, and random utility; they can be estimated within a Bayesian framework. Following Lauderdale and Clark (2014), we place Gaussian priors on the justice and case parameters:

$$\begin{aligned}\rho &\sim \mathcal{U}(0,1)\\ \boldsymbol{\psi}_j &\sim \mathcal{N}(\mathbf{0}, \lambda\mathbf{I} + \rho\mathbf{1}) \text{ for each justice } j\\ \boldsymbol{\kappa}_i^\top &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}) \text{ for each case } i\end{aligned}$$

The positive off-diagonal elements of the covariance matrix for justice IPs orient the issue-specific dimensions in the same direction (i.e with conservatives at the same end) and provide shrinkage of IP in each dimension to their common mean across dimensions. Fig. 1 presents the plate diagram for the IP models



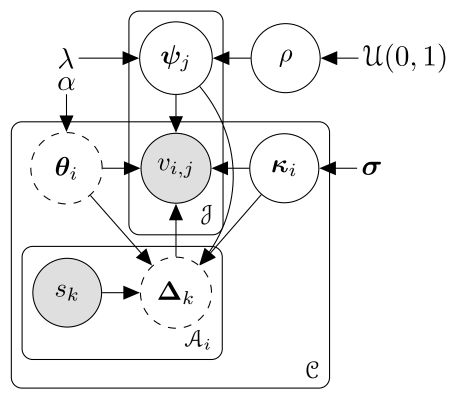


Figure 1: Plate diagrams for the IP models (left) and random utility model (right). $\mathcal{J}$, $\mathcal{C}$ and $\mathcal{A}_i$ are the sets of justices, cases, and amicus briefs (for case i), respectively. $\boldsymbol{\psi}_j$ is the IP for justice j; $\boldsymbol{\kappa}_i$ is the set of case parameters $a_i, b_i, c_i^{\mathsf{p}}$ and c_i^{r} for case i; and $\boldsymbol{\alpha}, \boldsymbol{\sigma}$, λ, and ρ are hyperparameters. The mixture proportion nodes (dashed) are fixed in our estimation procedure. On the left, black nodes comprise the basic IP model, blue nodes are found in the issues IP (§2.2) and amici IP (§2.3) models, while magenta nodes are found only in the amici IP model.

(§2.1, §2.2 and §2.3) on the left and the random utility model (§3) on the right.

For the non-utility IP models involving text (§2.2 and §2.3), LDA is used to infer the latent topic mixtures of text associated with the case and amicus briefs, thus

$$\boldsymbol{\theta}_i, \boldsymbol{\Delta}_k \sim \text{Dirichlet}(\alpha) \text{ for each case } i \text{ and brief } k$$

where the tokens in both sets of texts are drawn from shared topic-word distributions.[13]

[13]We omit details of LDA, as it is widely known.

Our random utility model can also be described through a similar generative story. Instead of drawing amicus briefs ($\boldsymbol{\Delta}$) from a Dirichlet, they are drawn from the expected utility distribution (Eq. 3). The right side of Fig. 1 shows the corresponding plate diagram. Importantly, note that $\boldsymbol{\Delta}$ here serves as direct evidence for the justice and case parameters, rather than influencing them through v-structures.

Parameter estimation. We decoupled the estimation of the topic mixture parameters as a stage separate from the IP parameters. This approach follows Lauderdale and Clark (2014), who argued for its conceptual simplicity: the text data defines the rotation of a multidimensional preference space, while the second stage estimates the locations in that space. We found in preliminary experiments that similar issue dimensions result from joint vs. stage-wise inference, but that the latter is much more computationally efficient.

Using LDA,[14] $\boldsymbol{\theta}$ (and, where relevant, $\boldsymbol{\Delta}$) are estimated, then fixed to their posterior means while solving for justice parameters $\boldsymbol{\psi}$ and case parameters $\boldsymbol{\kappa}$. For the second stage, we used Metropolis within Gibbs (Tierney, 1994), a hybrid MCMC algorithm, to sample the latent parameters from their posterior distributions. We sampled $\boldsymbol{\kappa}_i$ for each case and $\boldsymbol{\psi}_j$ for each justice blockwise from a multivariate Gaussian proposal distribution, tuning the diagonal covariance matrix to a target acceptance rate of between 15–45%. Likewise, ρ is sampled from a univariate Gaussian proposal, with its variance tuned similarly. For our random utility model, we used the same MCMC approach in sampling the latent IRT variables, but this time including the expected utility term for each brief in the likelihood function (Eq. 4). The details of our sampler can be found in the supplementary materials (§A).

Hyperparameters. We fixed the number of topics in our model to 30. For LDA, the symmetric Dirichlet parameter for document-topic and topic-word distributions are 0.1 and 0.001, respectively. For priors on the IP latent variables, we follow the same settings used by Lauderdale and Clark (2014), setting priors on case parameters, $\boldsymbol{\sigma}$, to 4.0, and justice IPs component-wise variance, λ to 1.0. In the random utility model, we set both hyperparameters η and ξ to 1.

5 Data Preprocessing

Cases	2,074
w/amicus briefs	1,531
Amicus briefs	7,258
Max. briefs/case	76
Word tokens	110.5M
Phrase tokens	8.9M
Ave. words/brief	3,094
Ave. phrases/brief	339

Table 1: Corpus statistics ignoring amicus briefs whose supporting side could not be automatically classified confidently. In the last row, briefs include merits and amicus briefs.

[14] We used the parallel C++ implementation of fast Gibbs sampling of LDA (Liu et al., 2011).

In our experiments, we focused on 23 terms of the Court from 1990–2012, obtaining cases and votes from the Supreme Court Database project (Spaeth et al., 2013),[15] and texts through LexisNexis' online subscription-based legal database.[16]

We concatenate each case's merits briefs from both parties to form a single document, where the text is used to infer the representation of the case in topical space ($\boldsymbol{\theta}$; i.e., merits briefs are treated as "facts of the case"). We did not make use of case opinions as did Lauderdale and Clark (2014) because opinions are written after votes are casted, tainting the data for predictive modeling. Each amicus brief is treated as a single document.

We tokenized the texts and tagged the tokens with the Stanford part of speech tagger (Toutanova et al., 2003). We extract n-grams with tags that follow the simple (but effective) pattern `(Adjective|Cardinal|Noun)+ Noun` (Justeson and Katz, 1995), representing each document as a "bag of phrases", and filtering phrases that appear in less than 25 or more than 3,000 documents, obtaining a vocabulary of 55,113 phrase types. Table 1 summarizes key details of our corpus.

Classifying amicus briefs. The amicus briefs in our dataset were not explicitly labeled with the side that they support, and manually labeling each brief would be a tedious endeavor. However, there are often cues in the brief content that *strongly* signal the side that the amici is supporting (e.g., "in support of petitioner" and "affirm the judgement"). Thus, we manually labeled 1,241 randomly selected amicus briefs with its side (petitioner, respondent, neither), and trained a classifier[17] using lexical and formatting features.[18]

We evaluated the performance of our classifier using 5 random splits, with 50% of our data for training, 30% for testing, and 20% for the development set.[19] The average accuracy of our classifier is 79.1%. Limiting our evaluation to instances whose posterior probability after classification is greater than 0.8, we obtain 90.0% accuracy and recall of 52.1%. Thus, we ignored 5,094 briefs (out of 13,162) that were classified as supporting neither side or whose posterior probability is ≤ 0.8 (higher precision at the expense of recall).

6 Experiments and Analysis

We trained our random utility model on the dataset and the topics and justices' IP are found in the supplementary materials (§B and §C, respectively).

6.1 Vote Prediction

We evaluate each model's ability to predict how justices would vote on a case out of the training sample. To compute the probability of justices' votes, we first infer the topic mixture proportions for the case's merits briefs ($\boldsymbol{\theta}$), and amicus briefs ($\boldsymbol{\Delta}$). Given all the justice's IPs $\boldsymbol{\psi}_j$, we find the most likely vote outcome for

[15] The unit of analysis is the case citation, and we select cases where the type of decision equals 1 (orally argued cases with signed opinions), 5 (cases with equally divided vote), 6 (orally argued per curiam cases), or 7 (judgements of the Court). In addition, we dropped cases where the winning side was not clear (i.e., coded as "favorable disposition for petitioning party unclear").

[16] `http://www.lexisnexis.com`

[17] We used a C++ implementation of logistic regression available at `https://github.com/redpony/creg`.

[18] We identified 5 sections which are common across almost all briefs and used them as features. The feature templates for our classifier are: $\langle w \rangle$, $\langle \text{title}, w \rangle$, $\langle \text{counsel}, w \rangle$, $\langle \text{introduction}, w \rangle$, $\langle \text{statement}, w \rangle$, $\langle \text{conclusion}, w \rangle$, where w can be any unigram, bigram, or trigram.

[19] We tuned the ℓ_1-regularization weights on our dev set over the range of coefficients $\{0.5, 1, 2, 4, 8, 16\}$.

the case by integrating over the case parameters $\boldsymbol{\kappa}$:[20]

$$\arg\max_{\boldsymbol{v}} \int_{\boldsymbol{\kappa}} p(\boldsymbol{\kappa} \mid \boldsymbol{\sigma}) \prod_{j \in \mathcal{J}} p(v_j \mid \boldsymbol{\psi}_j, \boldsymbol{\theta}, \boldsymbol{\Delta}, \boldsymbol{\kappa}) \\ \times \prod_{k \in \mathcal{A}} p_{\text{util}}(\boldsymbol{\Delta}_k \mid \boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{\Delta}, \boldsymbol{\kappa}, s_k)$$

where s_k is the side brief k supports, and the multiplier is the expected utility term (Eq. 3) which is ignored for the non-utility based models.

Due to the specification of IP models, we note that the the probability of a vote, which is a logistic function of the vote-specific IP, is a symmetric function. This implies that justice j's probability of voting towards the petitioner will be the same as if she voted for the respondent when we negate the vote-specific IP. Thus, we would not be able to distinguish the actual side that the justice will favor, but we can identify the most likely partitioning of the justices into the two groups.[21] We can then evaluate, for each case, the predictive ability by computing an average pairwise accuracy score,

$$\binom{9}{2}^{-1} \sum_{j,j' \in \mathcal{J}: j \neq j'} \mathbb{I}[\mathbb{I}[\hat{v}_j = \hat{v}_{j'}] = \mathbb{I}[v_j^* = v_{j'}^*]]$$

where $\hat{\boldsymbol{v}}$ ($\boldsymbol{v}^*$) are predicted (actual) votes.

We performed 5-fold cross validation, and present the votes partitioning accuracy in Table 2. A naïve

Model	**Accuracy**
Unanimous	0.714
Unidimensional IP (§2.1)	0.583
Issues IP (§2.2)	0.671
Amici IP (§2.3)	0.690
Random utility IP (§3)	**0.742**

Table 2: Accuracy of vote prediction.

baseline, where all justices vote unanimously, performs better than when adjusting for issues and/or amici. This suggests that justices' votes do not always align with their IP, and that topic models may be inadequate for representing IPs. Furthermore, we believe there may be insufficient information to learn the amicus polarity case parameters (c^s) in the amici IP model (which is slightly better than the issues IP model). However, in the random utility model, amici-agents-experts weigh in, and provide additional signals for estimating these parameters, achieving significant[22] predictive performance over the baseline.

6.2 Amicus Influence on Justices

Various justices have expressed opinions on the value of amicus briefs. Some justices, such as Scalia, are known to be dubious of amicus briefs, preferring to leave the task of reading these briefs to their law clerks, who will pick out any notable briefs for them.[23] In contrast, recent surveys of amicus citation rates in justice opinions have often ranked Sotomayor and Ginsburg among justices who most often cite amici in their opinions (Franze and Anderson, 2011; Walsh, 2013).

[20]We used the same MCMC technique described in §4.

[21]We expect that, given the partitioning of justices, domain experts would be able to identify the side each group of justices would favor.

[22]Paired samples t-test, $p < 0.001$.

[23]See *Roper v. Simmons*, 543 U.S. 551, 617-18 (2005) (Scalia J., dissenting); also, "Dont re-plow the ground that you expect the parties to plow unless you expect the parties to plow with a particularly dull plow" (attributed to Scalia on the California Appellate Law Blog, `http://bit.ly/1wOmlRE`).

For each justice, we compute the difference in vote log-likelihood between the issues IP and random utility IP models, as a (noisy) measure of amici influence. A larger difference in suggests that more of a justice's decision-making can be explained by the presence of amicus briefs. Table 3 presents these differences, and the rankings are consistent with extant hypotheses noted above. We take this consistency as encouraging, in the spirit of the "preregistered" hypotheses of Sim et al. (2013), but caution that it does not imply a conclusion about causation.

Justice	**Diff.** ($\times .01$)
C. Thomas	30.8
E. Kagan	28.9
S. Sotomayor	24.7
J. Roberts	24.0
R. Ginsburg	23.9
S. Breyer	22.2
S. Alito	21.8
A. Scalia	16.6
A. Kennedy	15.3

Table 3: Current justices ordered by how much amicus briefs are estimated to explain their votes, estimated as quadratic mean difference in vote log-likelihood between the issues IP model and the random utility IP model. The full list of justices can be found in §D of the supplementary material.

6.3 Post Hoc Analysis of Votes

On a case level, we can tease apart the relative contribution each textual component to a justice's decision by analyzing the case parameters learnt by our random utility model. By zeroing out various case parameters, and plotting them, we can visualize the different impact that each type of text has on a justice's vote-specific IP. For example, Fig. 2 shows the vote-specific IP estimates of justices for the 2011 term death penalty case *Maples v. Thomas*.[24] The issues-only IPs are computed by zeroing out both the amicus polarity parameters

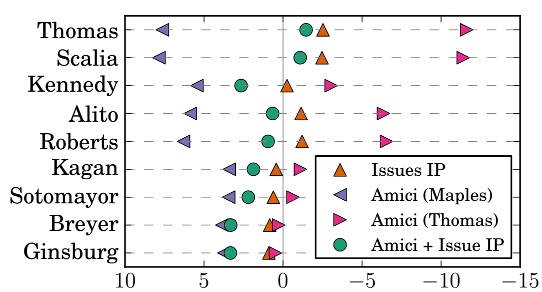


Figure 2: Vote-specific IP estimates decomposed into different influences on each justice's vote on *Maples v. Thomas*. An IP towards the left (right) indicates higher probability of vote that is favorable to Maples (Thomas).

(c^{p} and c^{r}). On the other hand, the IP due to amicus briefs supporting Maples (Thomas) is computed by zeroing out only c^{r} (c^{p}). We observe that the issues-only IPs are aligned with each justice's (widely known) ideological stance on the issue of capital punishment. For instance, the issues-only IPs of Thomas, Scalia,

[24] 132 S. Ct. 912 (2012).

Alito and Roberts, the strong conservative bloc, favor the respondents (that Maples should not be awarded relief); so did Kennedy, who is widely recognized as the swing justice. When the effects of all amicus briefs are taken into account, the justices' IPs shift toward Maples with varying magnitudes, with the result reflecting the actual ruling (7–2 with Thomas and Scalia dissenting).

6.4 Counterfactual Analysis

Following Pearl (2000), we can query the model and perform counterfactual analyses using the vote prediction algorithm (§6.1). As an illustration, we consider *National Federation of Independent Business (NFIB) v. Sebelius (HHS)*, a landmark 2011 case in which the Court upheld Congress's power to enact most provisions of the Affordable Care Act (ACA; "Obamacare").[25]

In the merits briefs, the topics discussed revolve around *interstate commerce* and the *individual mandate* (see Fig. 3 for the dominant topic proportions), while there is an interesting disparity in topics between briefs supporting NFIB and HHS. Notably, amici supporting NFIB are found, on average, to use language

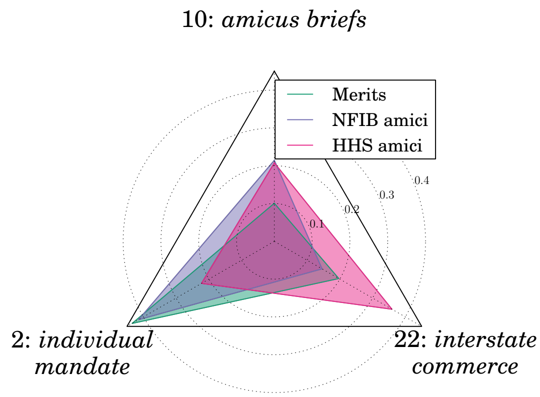


Figure 3: Top 3 topic proportions for merits briefs and amicus briefs. The *amicus briefs* topic is a collection of terms commonly associated with amicus briefs.

concerning *individual mandate*, while amici supporting HHS tend to focus more on topics related to *interstate commerce*. This is commensurate with the main arguments put forth by the litigants, where NFIB was concerned about the overreach of the government in imposing an individual mandate, while HHS argued that healthcare regulation by Congress falls under the Commerce Clause. During test time, our model was most uncertain about Roberts and Kennedy, and wrong about both (Fig. 5a).

Choosing sides. The first type of counterfactual analysis that we introduce is, "What if no (or only one side's) amicus briefs were submitted in the ACA case?" To answer it, we hold the case out of the training set and attempt to predict the votes under the hypothetical circumstances with the random utility model. Fig. 5a shows the resulting IP of hypothetical situations where no amicus briefs were filed, or when only briefs supporting one side are filed. If no amici filed briefs, the model expects that all but Kagan and Sotomayor would favor NFIB, but with uncertainty. With the inclusion of the amicus briefs supporting NFIB, the model becomes more confident that the conservative bloc of the court would vote in favor of NFIB (except for Alito). Interestingly, the model anticipates that the same briefs will turn the liberals *away*. In contrast, the

[25] 132 S. Ct. 2566 (2012). The case attracted much attention, including a record 136 amicus briefs, of which 76 of these briefs are used in our dataset. 58 (of the 76) were automatically classified as supporting NFIB.

briefs on HHS' side have more success in swaying the case in their favor, especially the crucial swing vote of Kennedy (although it turned out that Kennedy sided with the conservative bloc, and Roberts emerged as the deciding vote in HHS favor). Consequently, the model can provide insights about judicial decisions, while postulating different hypothetical situations.

Choosing what to write. Another counterfactual analysis we can perform, more useful from the viewpoint of the amicus, is, "how should an amicus frame arguments to best achieve her goals?" In the context of our model, such an amicus would like to choose the topic mixture $\boldsymbol{\Delta}$ to maximize her expected utility (Eq. 3). Ideally, one would compute such a topic mixture by maximizing over both $\boldsymbol{\Delta}$ and vote outcome $\boldsymbol{v}$, while integrating over the case parameters. We resort to a cheaper approximation: analyzing the filer's expected utility curve over two particular topic dimensions: the *individual mandate* and *interstate commerce* topics. That is, an amicus can choose between different proportions of the two topics to write about.[26]

Fig. 4 illustrates the expected utility curve faced by a single amicus as we vary the topic proportions of the *individual mandate* and *interstate commerce* topics. The model expects an amicus on NFIB's side to get more votes and hence, higher utility, as the model expects justices to be in favor of NFIB prior to amici influence (see Fig. 5a). Consequently, the amicus who supports NFIB can expect to maximize their expected utility (5.2 votes at a cost of 0.21) by "spending" about 70% of their text on *individual mandate*. On the other hand, the best that an amicus supporting HHS can do is to write a brief that is 80% about *interstate commerce*, and garner 4.7 votes at a cost of 0.31. We plot the justices' predicted IPs in Fig. 5b using these "best" proportions. The "best" proportions IPs are different (sometimes worse) from that in Fig. 5a because in Fig. 5a, there are multiple amici influencing the case parameters (through their utility functions) and other topics are present which will sway the justices. From the perspective of an amicus supporting HHS, the two closest swing votes in the case are Roberts and Kennedy; we know *a posteriori* that Roberts sided with HHS.

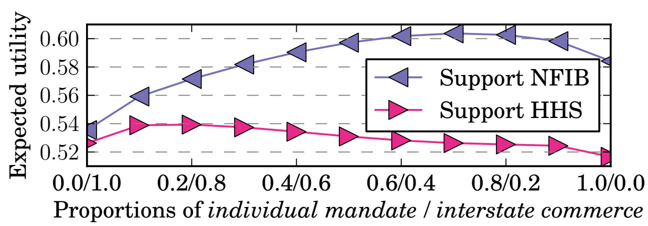
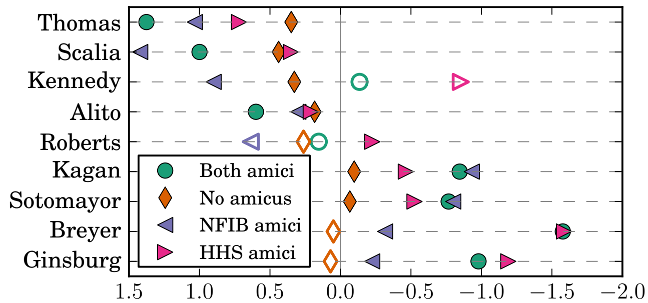


Figure 4: Expected utility when varying between proportions of *individual mandate* and *interstate commerce* topics.

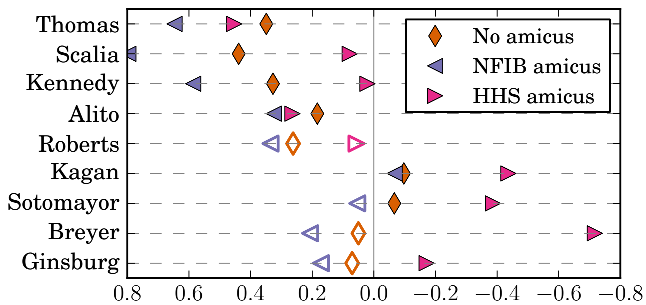
6.5 Discussion

Our model makes several simplifying assumptions: (i) it ignores the effects of other amici on a single amicus' writing; (ii) amici are treated modularly, with a multiplicative effect and no consideration of diminishing returns, temporal ordering, or reputation; (iii) the cost function does not capture the intricacies of legal writing style (i.e., choice of citations, artful language, etc.); (iv) the utility function does not fully capture the agenda of individual amicus; (v) each amicus brief is treated independently (i.e., no sharing across briefs with the same author), as we do not have access to clean author metadata. Despite these simplifications, the model is a useful tool for quantitative analysis and hypothesis generation in support of substantive research on the judiciary.

[26]Interior solutions for the expected utility optimization only exist when proportions are $\in (0, 1)$; we set the proportions of inactive topics to 10^{-8} instead of 0.



(a) What if amicus briefs for one side were not filed?



(b) What if a single amicus files an "optimally" written brief?

Figure 5: Counterfactual analyses for *National Federation of Independent Business v. Sebelius*. Hollow markers denote that the prediction differed from the actual outcome.

7 Related Work

Poole and Rosenthal (1985) introduced the IP model, using roll call data to infer latent positions of lawmakers. Since then, many varieties of IP models have been proposed for different voting scenarios: IP models for SCOTUS (Martin and Quinn, 2002), multidimensional IP models for Congressional voting (Clinton et al., 2004; Heckman and Snyder, 1996), grounding multidimensional IP models using topics learned from text of Congressional bills (Gerrish and Blei, 2012) and SCOTUS opinions (Lauderdale and Clark, 2014). Segal-Cover scores (Segal and Cover, 1989), obtained by manual coding of pre-confirmation news articles, are another popular method for characterizing behaviors of justices.

Amici have been studied extensively, especially their influence on SCOTUS (Caldeira and Wright, 1988; Collins, 2008; Corley et al., 2013; Kearney and Merrill, 2000). Particularly, Collins (2007) found that justices can be influenced by persuasive argumentation presented by organized interests. Hansford (2004) studied how amici decide whether to participate in a case, finding that amici participate in situations where justices are "information poor" or where cases allow for "high visibility." These studies focus on ideology metadata (liberal/conservative slant of amici, justices, decisions, etc.), disregarding the rich signals encoded in the text of these briefs, whereas we use the text as evidence of utility maximizing behavior to study the influence of amicus curiae.

Our model is also related to Gentzkow and Shapiro (2010) who model the purposeful "slant" of profit-maximizing newspapers looking to gain circulation from consumers with a preference for such slant. More generally, extensive literature in econometrics estimates structural utility-based decisions; for example Berry et al. (1995).

In addition to work on IP models, authorship (Li et al., 2013) and historical (Wang et al., 2012) analysis has been done on SCOTUS opinions, and oral argument transcripts have been used to study power relationships (Danescu-Niculescu-Mizil et al., 2012; Prabhakaran et al., 2013) and pragmatics (Goldwasser and Daumé III, 2014).

8 Conclusion

We have introduced a random utility model for persuasive text; it is similar to a classical generative model and can be estimated using familiar algorithms. The key distinction is that persuasive text is modeled as a function of the addressee and the particulars of the matter about which she is being convinced; authors are agents seeking to maximize their expected utility in a given scenario. In the domain of SCOTUS, this leads to improved vote prediction performance, as the model captures the structure of amicus briefs better than simpler treatments of the text. Secondly, and more importantly, our model is able to address interesting counterfactual questions. Were some amicus briefs not filed, or had they been written differently, or had the facts of the case been presented differently, or had different justices presided, our approach can estimate the resulting outcomes.

Acknowledgments

This research was supported in part by an A*STAR fellowship to Y. Sim, NSF CAREER grant IIS-1054319, a Google research award and supported by computing resources provided by a grant from the Pittsburgh Supercomputing Center.

References

Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890, 1995.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022, March 2003. ISSN 1532-4435. URL `http://dl.acm.org/citation.cfm?id=944919.944937`.

Gregory A. Caldeira and John R. Wright. Organized Interests and Agenda Setting in the U.S. Supreme Court. *American Political Science Review*, 82:1109–1127, December 1988. ISSN 1537-5943. doi: 10.2307/1961752. URL `http://journals.cambridge.org/article_S0003055400196352`.

Ming-Wei Chang, Lev Ratinov, and Dan Roth. Guiding semi-supervision with constraint-driven learning. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pages 280–287, June 2007. URL `http://cogcomp.cs.illinois.edu/papers/ChangRaRo07.pdf`.

Joshua Clinton, Simon Jackman, and Douglas Rivers. The Statistical Analysis of Roll Call Data. *American Political Science Review*, pages 355–370, May 2004. ISSN 1537-5943. doi: 10.1017/S0003055404001194. URL `http://journals.cambridge.org/article_S0003055404001194`.

Paul M. Collins. Lobbyists before the U.S. Supreme Court: Investigating the Influence of Amicus Curiae Briefs. *Political Research Quarterly*, 60(1):55–70, March 2007. doi: 10.1177/1065912906298535. URL `http://prq.sagepub.com/content/60/1/55.abstract`.

Paul M Collins. *Friends of the Supreme Court: Interest Groups and Judicial Decision Making*. Oxford University Press, August 2008. ISBN 019537214X. URL http://www.psci.unt.edu/~pmcollins/FOSC.htm.

Pamela Corley, Paul M Collins, and Jesse Hamner. The Influence of Amicus Curiae Briefs on US Supreme Court Opinion Content. In *American Political Science Association 2013 Annual Meeting Paper*, August 2013.

Cristian Danescu-Niculescu-Mizil, Lillian Lee, Bo Pang, and Jon Kleinberg. Echoes of power: Language effects and power differences in social interaction. In *Proceedings of WWW*, pages 699–708, 2012.

Jean-Paul Fox. *Bayesian Item Response Modeling: Theory and Applications*. Statistics for Social and Behavioral Sciences. Springer, June 2010.

Anthony J. Franze and R. Reeves Anderson. Commentary: The Courts increasing reliance on amicus curiae in the past term. http://www.arnoldporter.com/resources/documents/Arnold&PorterLLP_NationalLawJournal_8.24.11.pdf, August 2011.

Kuzman Ganchev, João Graça, Jennifer Gillenwater, and Ben Taskar. Posterior regularization for structured latent variable models. *Journal of Machine Learning Research*, 11:2001–2049, August 2010. ISSN 1532-4435. URL http://dl.acm.org/citation.cfm?id=1756006.1859918.

Matthew Gentzkow and Jesse M Shapiro. What drives media slant? Evidence from US daily newspapers. *Econometrica*, 78(1):35–71, 2010.

Sean Gerrish and David M. Blei. How They Vote: Issue-Adjusted Models of Legislative Behavior. In *Advances in Neural Information Processing Systems 25*, pages 2753–2761, December 2012.

Dan Goldwasser and Hal Daumé III. "I Object!" Modeling Latent Pragmatic Effects in Courtroom Dialogues. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 655–663, Gothenburg, Sweden, April 2014.

Thomas G. Hansford. Information Provision, Organizational Constraints, and the Decision to Submit an Amicus Curiae Brief in a U.S. Supreme Court Case. *Political Research Quarterly*, 57(2):219–230, 2004.

James J. Heckman and James M. Snyder, Jr. Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators. Working Paper 5785, National Bureau of Economic Research, October 1996. URL http://www.nber.org/papers/w5785.

Geoffrey E Hinton. Training Products of Experts by Minimizing Contrastive Divergence. *Neural Computation*, 14(8):1771–1800, August 2002. ISSN 0899-7667. URL http://dx.doi.org/10.1162/089976602760128018.

Simon Jackman. Multidimensional Analysis of Roll Call Data via Bayesian Simulation: Identification, Estimation, Inference, and Model Checking. *Political Analysis*, 9(3):227–241, June 2001. doi: 10.1093/polana/9.3.227. URL http://pan.oxfordjournals.org/content/9/3/227.abstract.

John S. Justeson and Slava M. Katz. Technical Terminology: Some Linguistic Properties and an Algorithm for Identification in Text. *Natural Language Engineering*, 1:9–27, March 1995.

Joseph D Kearney and Thomas W Merrill. The influence of amicus curiae briefs on the supreme court. *University of Pennsylvania Law Review*, pages 743–855, 2000.

Benjamin E. Lauderdale and Tom S. Clark. Scaling Politically Meaningful Dimensions Using Texts and Votes. *American Journal of Political Science*, 2014. ISSN 1540-5907. doi: 10.1111/ajps.12085. URL http://dx.doi.org/10.1111/ajps.12085. To appear.

William Li, Pablo Azar, David Larochelle, Phil Hill, James Cox, Robert C Berwick, and Andrew W Lo. Using algorithmic attribution techniques to determine authorship in unsigned judicial opinions. *Stanford Technology Law Review*, pages 503–534, 2013.

Zhiyuan Liu, Yuzhou Zhang, Edward Y. Chang, and Maosong Sun. PLDA+: Parallel Latent Dirichlet Allocation with Data Placement and Pipeline Processing. *ACM Transactions on Intelligent Systems and Technology, Special issue on Large Scale Machine Learning*, 2011. Software available at `http://code.google.com/p/plda`.

Kelly J Lynch. Best Friends - Supreme Court Law Clerks on Effective Amicus Curiae Briefs. *Journal of Law & Politics*, 20:33, Winter 2004.

Andrew D. Martin and Kevin M. Quinn. Estimating Latent Structures of Voting for Micro-Committees with Application to the U.S. Supreme Court. In *Midwest Political Science Association 2001 Annual Meeting Paper*, 2001.

Andrew D Martin and Kevin M Quinn. Dynamic ideal point estimation via Markov chain Monte Carlo for the US Supreme Court, 1953-1999. *Political Analysis*, 10(2):134–153, May 2002.

Andrew McCallum, Gideon Mann, and Gregory Druck. Generalized Expectation Criteria. Technical Report UM-CS-2007-60, University of Massachusetts, Amherst, MA 01003, USA, August 2007.

Daniel McFadden. Conditional logit analysis of qualitative choice behavior. In Paul Zarembka, editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York, 1974.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000.

Keith T. Poole and Howard Rosenthal. A Spatial Model for Legislative Roll Call Analysis. *American Journal of Political Science*, 29(2):357–384, May 1985. ISSN 00925853. URL `http://www.jstor.org/stable/2111172`.

Vinodkumar Prabhakaran, Ajita John, and Dorée D. Seligmann. Who had the upper hand? ranking participants of interactions based on their relative power. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 365–373, Nagoya, Japan, October 2013. Asian Federation of Natural Language Processing.

Jeffrey A. Segal and Albert D. Cover. Ideological Values and the Votes of U.S. Supreme Court Justices. *The American Political Science Review*, 83(2):557–565, June 1989. ISSN 00030554. URL `http://www.jstor.org/stable/1962405`.

Yanchuan Sim, Brice D. L. Acree, Justin H. Gross, and Noah A. Smith. Measuring ideological proportions in political speeches. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 91–101, Seattle, WA, USA, October 2013.

Harold J. Spaeth, Sara Benesh, Lee Epstein, Andrew D. Martin, Jeffrey A. Segal, and Theodore J. Ruger. Supreme Court Database, Version 2013 Release 01, 2013. URL `http://supremecourtdatabase.org`. Data available at `http://supremecourtdatabase.org`.

Luke Tierney. Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics*, 22(4):1701–1728, December 1994. ISSN 00905364. URL `http://www.jstor.org/stable/2242477`.

Kristina Toutanova, Dan Klein, Christopher D. Manning, and Yoram Singer. Feature-rich Part-of-speech Tagging with a Cyclic Dependency Network. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, pages 173–180, Edmonton, Canada, 2003.

Mark Walsh. Supreme Court Report: It was another big term for amicus curiae briefs at the high court. `http://www.abajournal.com/magazine/article/it_was_another_big_term_for_amicus_curiae_briefs_at_the_high_court/`, September 2013.

William Yang Wang, Elijah Mayfield, Suresh Naidu, and Jeremiah Dittmar. Historical analysis of legal opinions with a sparse mixed-effects latent variable model. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 740–749, Jeju Island, Korea, 2012.

Appendices

A Likelihood of Random Utility Ideal Point Model

The likelihood of our amici model with random utility is given below,

$$\begin{aligned}
&p(\boldsymbol{v}, \rho, \boldsymbol{\psi}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}^{\mathrm{p,\,r}} \mid \lambda, \boldsymbol{\theta}, \boldsymbol{\Delta}, \boldsymbol{s}) \\
&\propto \prod_{j \in \mathcal{J}} p(\boldsymbol{\psi}_j \mid \lambda, \rho) \prod_{i \in \mathcal{C}} p(a_i, b_i, c_i^{\mathrm{p,\,r}} \mid \boldsymbol{\sigma}) \\
&\times \prod_{i \in \mathcal{C}} \prod_{j \in \mathcal{J}} V_{i,j} \prod_{k \in \mathcal{A}_i} \mathbb{E}[U_{i,k}]^{\eta}
\end{aligned}$$

where

$$\begin{aligned}
V_{i,j} &= p(v_{i,j} \mid \boldsymbol{\psi}_j, \boldsymbol{\theta}_i, \boldsymbol{\Delta}_i, a_i, b_i, c_i^{\mathrm{p,\,r}}) \\
&= \sigma \left[(1 - v_{i,j}) a_i \right. \\
&\quad \left. + (1 - v_{i,j}) \boldsymbol{\psi}_j^{\top} (b_i \boldsymbol{\theta}_i + c_i^{\mathrm{p}} \boldsymbol{\Delta}_i^{\mathrm{p}} + c_i^{\mathrm{r}} \boldsymbol{\Delta}_i^{\mathrm{r}}) \right],
\end{aligned}$$

$\sigma(x) = \frac{\exp(x)}{1+\exp(x)}$ is the logistic function, and

$$\begin{aligned}
\mathbb{E}[U_{i,k}] = &\sum_{j \in \mathcal{J}} \sigma \left(a_i + \boldsymbol{\psi}_j^{\top} (b_i \boldsymbol{\theta}_i + c_i^{s_{i,k}} \boldsymbol{\Delta}_{i,k}) \right) \\
&+ \xi \left(1 - \frac{1}{2} \|\boldsymbol{\Delta}_{i,k} - \boldsymbol{\theta}_i\|_2^2 \right)
\end{aligned}$$

We add a constant to the expected utility term so that it will always be ≥ 0. During each iteration of Gibbs sampling, we sampled each latent variable $\rho, \boldsymbol{\psi}_j$ and $[a_i, b_i, c_i^{\mathrm{p,\,r}}]$ blockwise from the likelihood in turn using the Metropolis-Hastings algorithm. In each Metropolis-Hastings random walk, we took 500 steps, ignoring the first 250 for burn-in and keeping every 10th step to compute the mean, which we use as our Gibbs update. In total, we performed 2,000 Gibbs iterations over the training data.

B Topic Distribution

Table 5 lists the topics and top phrases estimated from our dataset using LDA.

C Justices' Ideal Points

The ideal points of justices vary depending on the issues. We present the justices' ideal points for each of the 30 topics in Fig. 6.

D Amici Influence on Justices

Table 4 lists justices in our dataset ordered by the difference of their vote log-likelihood under the issues IP model and random utility IP model.

Justice	**Diff.** $(\times .01)$
John Paul Stevens	31.5
Clarence Thomas*	30.8
Thurgood Marshall	29.4
Elena Kagan*	28.9
Sonia Sotomayor*	24.7
John Roberts*	24.0
Ruth Bader Ginsburg*	23.9
William Rehnquist	23.8
Stephen Breyer*	22.2
David Souter	22.0
Samuel Alito*	21.8
Harry Blackmun	18.0
Byron White	16.9
Antonin Scalia*	16.6
Anthony Kennedy*	15.3
Sandra Day O'Connor	14.7

Table 4: Justices ordered by how much amicus briefs can explain their votes, estimated as quadratic mean difference in vote log-likelihood between the issues IP model and the random utility IP model. *Active justices.

#	Topic	Top phrases
1	Criminal procedure (1)	reasonable doubt, supervised release, grand jury, prior conviction, plea agreement, controlled substance, guilty plea, double jeopardy clause, sixth amendment, jury trial
2	Employment	erisa plan, plan administrator, employee benefit plan, insurance company, pension plan, health care, plan participant, individual mandate, fiduciary duty, health insurance
3	Due process	due process clause, equal protection clause, fundamental right, domestic violence, equal protection, state interest, d e, assisted suicide, controlled substance, rational basis
4	Indians	m r, m s, indian tribe, tribal court, indian country, fifth amendment, miranda warning, indian affair, vice president, tribal member
5	Economic activity	attorney fee, limitation period, hobbs act, security law, rule 10b, actual damage, racketeering activity, fiduciary duty, loss causation, security exchange act
6	Bankruptcy law	bankruptcy court, bankruptcy code, 1996 act, state commission, telecommunication service, network element, eighth circuit, new entrant, pole attachment, communication act
7	Voting rights	voting right, minority voter, j app, voting right act, covered jurisdiction, fifteenth amendment, redistricting plan, political process, political subdivision, minority group
8	First amendment	first amendment right, commercial speech, strict scrutiny, cable operator, free speech, first amendment protection, protected speech, child pornography, government interest, public forum
9	Taxation	interstate commerce, commerce clause, state tax, tax court, gross income, internal revenue code, income tax, dormant commerce clause, state taxation, sale tax
10	Amicus briefs	national association, amicus brief, vast majority, brief amicus curia, large number, wide range, recent year, public policy, wide variety, washington dc
11	Labor management	north carolina, collective bargaining agreement, confrontation clause, sta t, north platte river, collective bargaining, inland lake, laramie river, labor organization, re v
12	Civil action	class action, class member, injunctive relief, final judgment, federal claim, civil action, preliminary injunction, class certification, civil procedure, subject matter jurisdiction
13	Civil rights	title vii, title vi, civil right act, age discrimination, sexual harassment, old worker, major life activity, reasonable accommodation, prima facie case, disparate impact
14	State sovereign	sovereign immunity, eleventh amendment, state official, absolute immunity, false claim, private party, 42 usc 1983, state sovereign immunity, eleventh amendment immunity, federal employee
15	Federal administrations	federal agency, statutory construction, plain meaning, other provision, statutory text, dc circuit, sub (a), fiscal year, senate report, agency action
16	Interstate relations	special master, new mexico, prejudgment interest, arkansas river, rt vol, comp act, new jersey, elli island, john martin reservoir, video game
17	Court of Appeals	eleventh circuit, sixth circuit, circuit court, fourth circuit, oral argument, tenth circuit, further proceeding, appeal decision, instant case, defendant motion
18	Fourth amendment	fourth amendment, probable cause, arbitration agreement, police officer, national bank, search warrant, exclusionary rule, arbitration clause, reasonable suspicion, law enforcement officer
19	Eighth amendment	eighth amendment, sex offender, prison official, facto clause, copyright act, copyright owner, unusual punishment, public domain, liberty interest, public safety
20	International law	international law, foreign state, vienna convention, human right, foreign country, foreign government, jones act, united kingdom, native hawaiian, foreign nation
21	Equal protection clause	peremptory challenge, law school, equal protection clause, strict scrutiny, high education, racial discrimination, prima facie case, school district, consent decree, compelling interest
22	Commerce clause (2)	interstate commerce, commerce clause, local government, political subdivision, state regulation, supremacy clause, federal regulation, tobacco product, tenth amendment, federal fund
23	Immigration law	judicial review, immigration law, final order, removal proceeding, immigration judge, due process clause, deportation proceeding, administrative remedy, compliance order, time limit
24	Death penalty	death penalty, habeas corpus, reasonable doubt, trial judge, death sentence, ineffective assistance, direct appeal, defense counsel, new rule, mitigating evidence
25	Environmental issues	navigable water, clean water act, colorado river, special master, project act, public land, fill material, water right, point source, lake mead
26	Establishment clause	establishment clause, school district, public school, private school, religious school, ten commandment, boy scout, religious belief, religious organization, free exercise clause
27	Patent law	federal circuit, patent law, prior art, subject matter, expert testimony, lanham act, hazardous substance, patent system, patent act, new drug
28	Antitrust law	antitrust law, sherman act, contr act, market power, postal service, joint venture, natural gas, high price, public utility, interstate commerce
29	Election law	political party, taking clause, private property, property owner, fifth amendment, independent expenditure, federal election, property right, contribution limit, general election
30	Criminal procedure (2)	punitive damage, habeas corpus, second amendment, punitive damage award, enemy combatant, military commission, compensatory damage, state farm, new trial, due process clause

Table 5: Topics and top-10 phrases estimated from briefs using LDA. We manually annotated each topic with the topic labels.

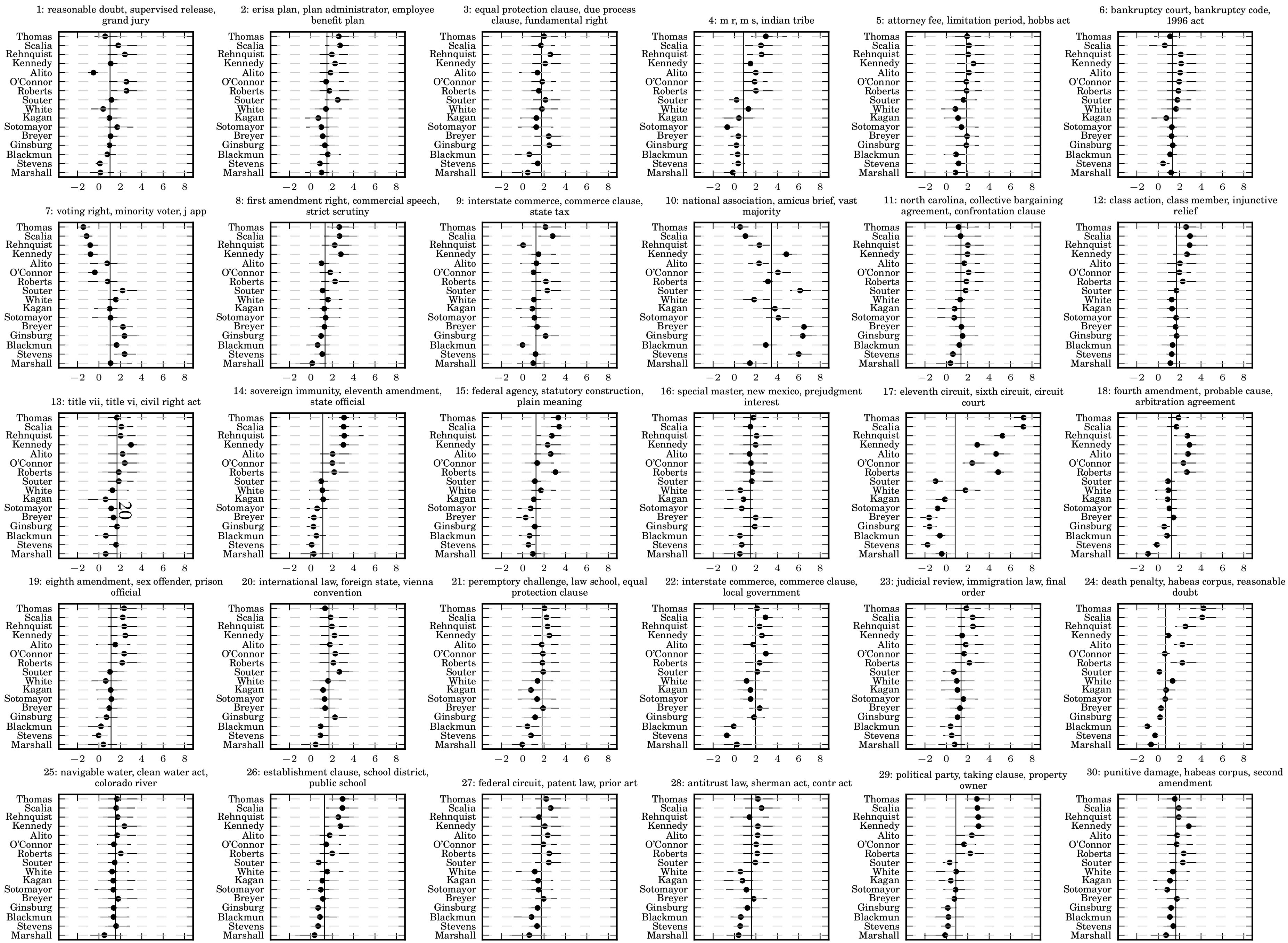


Figure 6: Justices' ideal points by topics. Solid vertical line denotes median ideal point of the justices.