Measuring the Issue Content of Supreme Court Opinions

DOUGLAS RICE, University of Massachusetts Amherst

ABSTRACT
The opinions of the US Supreme Court are central to volumes of research on law, courts, and politics. To understand these complex and often-lengthy documents, scholars frequently rely on dichotomous indicators of opinion content. While sometimes appropriate, for many research settings this simplification of opinion content systematically omits important information. Using all US Supreme Court opinions from 1803 to 2010 in association with structural topic models, I instead demonstrate the value of representing the Court’s attention in opinions in terms of topic proportions.

Perched atop the judiciary, the US Supreme Court regularly publishes opinions that provide critical information on law and public policy. Understanding the information contained within opinions has therefore served as a principal objective for social scientists and legal scholars for generations. As is frequently recognized, the identification of this dimension is critical to our understanding of the Court, including for example debate over strategic judicial behavior (Friedman 2006; Benesh and Spaeth 2007) and the Court’s policy-making role (Baird 2004, 2007; Peters 2007).

Yet identifying and measuring the content of a judicial opinion has proven remarkably controversial, as parsing voluminous and verbose documents poses a particular challenge for the researcher. At the heart of this challenge is what I term the dimensionality problem. Even within a case, the Court is often called to at least consider if not resolve multiple issues. Simply at a minimum, questions over justiciability are addressed in all cases reaching the Court in addition to whatever specific controversy is at issue. Take, for example, the landmark abortion decision Roe v. Wade; beyond establishing a trimester framework for evaluating the constitutionality of abortion restrictions, the opinion addressed standing

I would like to thank Tom Clark, Marie Hojnacki, James Honaker, Burt Monroe, Joseph Smith, and Chris Zorn for their helpful comments and suggestions on prior iterations of this research. Data and supporting materials necessary to reproduce the numerical results in the article are available in the JLC Dataverse at https://doi.org/10.7910/DVN/YFLR9W. Contact the author at drice@umass.edu.
and justiciability questions at length. For the researcher, incorporating each additional dimension may yield a better and more accurate understanding of the Court’s attention, but it correspondingly complicates the classification task. Moreover, even if one identified the presence of multiple dimensions, the measurement would still obscure the relative attention the Court paid to each of those dimensions. Given the challenges of identifying multiple dimensions, along with the relative impossibility of identifying attention proportionally through standard human-coding approaches, scholars regularly resort to single membership classification approaches, relying on measures that identify only a single, primary dimension.

For this, scholars almost exclusively rely on the high-quality issue coding available from the Supreme Court Database (SCD). Explicitly intended to measure the policy content (or “substantive basis”) of the opinion, the coding protocol places a strong preference on assigning only a single issue for each case. In so doing, the codes reduce lengthy treatises to—in the overwhelming majority of cases—a single issue dimension that was the basis of the Court’s decision. Given their high quality, accessibility, and widespread adoption, the codes have proven valuable in enhancing the comparability of a long history of research.

Yet the cost of the simplification is the potential loss of information. The preference for assigning a single membership classification potentially obscures large amounts of important information for scholars interested in the Court. While the loss may be acceptable for some research applications, it limits the opportunities for other research to fully understand the Court’s attention and judicial behavior. Alternatively, the classification of issues—or topics—is precisely the sort of task for which computational approaches to text analysis are ideally suited. In this article, I demonstrate the utility of one such computational method—structural topic models (Roberts et al. 2014)—as a flexible measurement approach offering estimates of topic attention across multiple dimensions. I estimate the Court’s topical attention across a corpus of all Supreme Court majority opinions between 1803 and 2010. Through a comparison of the estimated topic proportions to the SCD issue variables both for specific cases and as measures of changes in the Court’s agenda, I document the validity of the estimated topics. Then, I offer a series of illustrations for areas in which the estimated topic proportions provide an opportunity to address substantively important questions for which dichotomous indicators are potentially insufficient. Finally, I conclude by discussing how the flexibility of the approach offers a wealth of opportunities for extensions in subsequent research.

1. Indeed, Epstein, Knight, and Martin (2003, 812), in their review of research on judicial behavior, identify the SCD as “so dominating in our discipline that it would certainly be unusual for a refereed journal to publish a manuscript whose data derived from an alternative source.”

2. A series of preestimated models of opinion content across different levels of specificity is available in the JLC Dataverse at https://doi.org/10.7910/DVN/VFXPQG.
MEASURING OPINION CONTENT

Identifying the issues addressed by a given court case is critically important for research on law and courts. Often, scholars restrict their analyses to particular areas of the law in order to better identify a specific effect, say of case facts or precedent, on judicial behavior (e.g., Segal 1984, 1986; Richards and Kritzer 2002; Richards, Smith, and Kritzer 2006). In yet other work, scholars frequently rely on model specifications that include dichotomous covariates (fixed effects) for issue areas, seeking to address issue-specific heterogeneity. Likewise, specifying the issue at debate serves as the principal means to identify the Court’s ideological output (e.g., McGuire et al. 2009) and to identify dimensions of dispute in the case (e.g., Benesh and Spaeth 2007). Even further, the identification of issues has implications for the rhetorical strategies of the justices, including whether they engage in heresitical maneuvering (e.g., Rice 2016) or issue creation (e.g., McAtee and McGuire 2007). Moving beyond judicial behavior, measuring issue content also has important implications for cross-institutional studies, including understanding the judicial hierarchy (e.g., Baird 2004, 2007; Peters 2007), legal mobilization (e.g., Rice 2014), and the influence of the courts in the policy process (e.g., Flemming, Bohte, and Wood 1997; Flemming, Wood, and Bohte 1999). In sum, an understanding of the issues addressed by the Court is implicated in nearly all areas of research on the Court.

To that end, the identification of the issue continues to generate discussion and disagreement among top scholars in the field, including an August 2016 exchange on the public discussion list of the Law and Courts Section of the American Political Science Association.3 In the first place, there is disagreement as to the very definition of issue. The SCD, for instance, classifies issues based on public policy or the “subject matter of the controversy.” Even within this definition, however, there are conceptual difficulties, as categories such as “judicial power” are generally not perceived to be public policy issues.4 In another project, Shapiro (2009) classifies instead legal issues that transition away from a public policy focus and are instead explained as “imagin[ing] that you are trying to describe the case to a first-year law student. You want the student to understand what the case is about and what areas of law it implicates” (532). Here, I build on these definitions and consider issues or topics thematically. That is, I consider issues as the themes that characterize a corpus of text. This definition has the benefit of making plain that which is implicit in both of the above-mentioned definitions: there are multiple potential nestings and hierarchies of topics, such that any division ultimately rests on the substantive interests of the scholars employing the measures. Moreover, this definition does not require differentiating policy from legal issues. Instead, both can be captured at least insofar as they represent important variation in themes across the corpus.

4. Note, e.g., that the Policy Agendas Project does not include any references to judicial power in their topic or subtopic categories.
The primary difficulty of the manual coding enterprise, though, lies in the dimensionality problem. Court opinions are lengthy, complex documents and therefore regularly address multiple issues. Moreover, emphasizing or prioritizing one of those issues over the others may often be a key component of the Court’s decision-making process (Epstein and Shvetsova 2002; Wedeking 2010; Black, Schutte, and Johnson 2013; Rice 2016). Yet identifying each issue vastly complicates the coding task, requiring detailed analysis of the entire opinion. Moreover, there is no clear stopping rule for the coder; that is, what constitutes enough attention? Is a footnote adequate? Both of these problems are then magnified in that this dichotomous treatment would treat issues that are discussed in brief equivalently to those issues that formed the primary basis of the Court’s decision. Finally, the proliferation of these dimensions increases the probability of discrepancies in coding decisions, as the reliability of measures decreases as the complexity of the task increases.

To overcome these challenges, researchers reasonably rely on publicly available issue measures from the SCD, which provides human-coded public policy and legal issue codes. For each case, the classification scheme addresses the dimensionality problem by explicitly identifying the single most important dimension in the Court’s decision. That is, the SCD treats opinions as a single membership classification; therefore, every one of the Court’s opinions is classified into one of 260 unique codes further organized into one of 14 broad categories called issue areas.5 These codes are quite explicitly intended to reflect the public policy content of the opinion; as the coding protocol states, “this variable identifies issues on the basis of the Court’s own statements as to what the case is about. The objective is to categorize the case from a public policy standpoint” (42). Thus, for nearly every opinion of the Court, the issue codes identify a single policy that the Court (justice) addressed in that opinion.

In relying on these issue codes, researchers reasonably resolve three of the major challenges in identifying issue content. First, the public availability of high-quality, consistent, and reliable coding of issue attention is available for free and for all Supreme Court decisions. Thus, researchers are able to mitigate the massive resource costs necessary to identify these codes themselves. Second, by identifying a single, primary dimension for the decision, researchers explicitly avoid the equivalency problem. The coding protocol specifically advocates identifying the primary dimension on which the decision was made, and thus all subordinate issues—although potentially—are omitted. Finally, the prevalence of the codes—and the prevalence of the SCD more generally—enhances the comparability of research. That is, while two separate researchers may come to different conclusions as to the primary issue of a case, relying on a single estimate across different research settings enhances our ability to compare identified relationships.

Yet the approach carries an important cost. By relying on single membership classifications, the measure necessarily simplifies complex legal documents. Importantly, in many

---

5. On rare occasions, the SCD does code multiple topics for a single case. However, these instances are exceedingly rare and arise only when the coder cannot identify a single primary dimension.
cases such an approach likely leads to a large-scale deletion of potentially important information. Were cases generally about a single issue, the deletion may be ignorable. But the presence of multiple issues in a case is well recognized, as minimally the Court must address justiciability concerns in addition to whatever controversy it is being asked to decide. Consider the SCD’s classification of Wiggins v. People in Utah (1876), which is identified as disposed on the grounds of judicial power, specifically jurisdiction. The entirety of the majority opinion’s discussion of jurisdiction is as follows: “Sec. 3 of the Act of Congress of June 23, 1874, 18 Stat. 254, allows a writ of error from this Court to the Supreme Court of the Territory of Utah, where the defendant has been convicted of bigamy or polygamy or has been sentenced to death for any crime. The present writ is brought under that statute to obtain a review of a sentence of death against plaintiff in error for the murder of John Kramer, commonly called Dutch John, in Salt Lake City” (93 U.S. at 466). The remaining six pages of the majority opinion, and the entirety of Justice Clifford’s dissenting opinion, focused on a description of the homicide at the heart of the case and the admissibility of witness testimony to the jury. Thus, although the Court certainly considered and was required to address the issue of jurisdiction—as they must in any case—the overwhelming focus of Wiggins was on a separate issue that the Court also addressed.

To understand the scale of the potential deletion, Shapiro (2009) revisited a random sample of 95 opinions and identified each of the issues addressed in the opinions. In this context, Shapiro used the SCD’s coding protocol but did not give preference to assigning only a single issue. Of the 95 cases recoded, Shapiro identified 89—or approximately 94% of the opinions—as addressing more than one issue. Moreover, for many of those cases Shapiro identified more than two issues in the case. In contrast, only one of those 89 opinions was classified as addressing multiple issues by the SCD.

Considering the implications of deleting information for statistical inference (see, e.g., Gelman and Hill 2007), the implications of the unidimensional assumption are potentially profound for some research settings. Particularly as researchers seek to better identify and understand judicial behavior—to “take law seriously” (Friedman 2006)—generating and using appropriate measures of opinion content are essential. In the remainder of this article, I turn to this task.

TOPIC MODELS FOR MEASURING OPINION CONTENT

I do so by employing tools from a well-developed area of computational social science. The potential utility of computational approaches to studies of the Supreme Court and other judicial institutions is well demonstrated by the volume of research in the area. In the first place, researchers have looked to improve variables already included in the SCD (e.g., McGuire and Vanberg 2005; Evans et al. 2007). For example, McGuire and Vanberg (2005) use Wordscores (Laver, Benoit, and Garry 2003) to measure the ideological direction of Supreme Court opinions, an approach built upon by Lauderdale and Clark (2014), who jointly model votes and issues to extract ideological dimensions within issue areas.
In addition to refining existing measures, scholars have proposed expanding the coding scheme to other judicial venues (Szmer and Edwards 2011).

In this vein, the unsupervised estimation of topical content from text offers an avenue to resolve the underlying problems of the dimensionality problem in a flexible and cost-efficient way. Topic models are a suite of methods for organizing and understanding collections of documents. In this vein, topics can be understood as clusters of co-occurring words. This clustering of co-occurring words fits well with the definition of topics as themes that define the corpus. Within political science alone, increasingly sophisticated research has been conducted on topic models to perform unsupervised topic classification (Grimmer 2010; Quinn et al. 2010) with recent efforts at creating more flexible topic modeling forms that can adopt additional information (Roberts, Stewart, and Airoldi 2016). Models uncover patterns of word use across documents and connect those documents that exhibit similar patterns (Blei and Lafferty 2009). Note that, unlike human-coded schemes, topics are explicitly determined by the language of the documents. Subsequent analysis within one of these topic areas would thus be able to differentiate other dimensions of interest within an issue area, such as ideological direction (Lauderdale and Clark 2014).

Multiple algorithms are available for unsupervised topic classification. The standard, vanilla approach is latent Dirichlet allocation, or LDA. In LDA and many associated topic modeling approaches, documents are considered as arising from multiple latent topics—that is, documents can belong to multiple classifications or mixed membership classification—an important improvement over many other clustering algorithms which classified documents into single topics, that is, single membership classification (Blei, Ng, and Jordan 2003). Assuming a certain number of topics associated with a set of documents and having observed the words that make up the documents, we estimate the proportion of each document that exhibits each topic (Blei and Lafferty 2009). The primary value of the approach lies in the estimation of the relative proportion of each document arising from a topic, or topic proportions. In so doing, the approach allows researchers to move past unidimensional assumptions to more granular reflections of the Court’s attention.

One major advance in recent years was the introduction of structural topic models (STMs) by Roberts et al. (2013). STMs advance prior work by allowing users to incorporate document-level covariates affecting the estimation of topics. For instance, one common covariate affecting the estimation of topics is time; in our context, the Supreme Court is well recognized as having an agenda that shifted markedly over time (e.g., Pacelle 1995). By incorporating document-level covariates for the term of the Court, one can incorporate our understanding of factors affecting agenda change while also directly and more appropriately modeling the Court’s issue attention. Applications of the STM modeling ap-

---

6. Inference is performed on the posterior probabilities through one of multiple possible methods (Blei et al. 2003), frequently Gibbs sampling (Griffiths and Steyvers 2004).
proach have already yielded important insights for a diversity of relevant political science work, including survey research (Roberts et al. 2014), international bargaining (Tingley 2017), lobbying and trade (Kim 2017), and constitutional design (Law 2016). In this vein, I estimate a series of STMs to demonstrate the value of this approach, and mixed membership topic models more generally, for understanding the issue attention of the Court.

ESTIMATING TOPIC ATTENTION

To begin, I acquired the texts of all majority opinions in the Court’s history from Justia, an online repository of legal information. In total, the corpus features 25,291 majority opinions, with the earliest opinions in 1803 and the most recent opinions in 2010. I retain the 5,000 most frequent unigrams (i.e., “constitution,” “congress”) and the 2,500 most frequent bigrams (i.e., “civil rights,” “congressional power”). I include bigrams in order to capture particular legal concepts that may be muddled in a unigram-only approach.

Although STM requires very few predefined parameters, it does require prior selection of the number of topics (k) before estimation. Selection of the number of topics is, as Blei and Lafferty (2009) state, “a persistent problem in topic modeling” (11) as there exists no widely agreed method for selecting the “correct” value. The problem arises, however, from the flexibility of the model; researchers can adjust the parameter in order to obtain more general or more specific estimates of topical content, and for different research enterprises different values are potentially appropriate. Rather than settling on a single k here, I instead extract latent topic proportions across multiple specifications in the analyses to follow. Finally, I create topic names by concatenating the three most representative terms by topic.

Measurement models are assessed by a number of validity criteria. As evidence of the face validity of the measure as well as convergent (e.g., does the measure map to other measures that it should map to) and discriminant (e.g., does the measure diverge from other measures in useful ways) validity, I look here to topics identified through a 14-topic STM. I use 14 topics in order to maximize comparability to the 14 issue areas identified by the SCD. In table 1, I present the top three words characterizing each of the estimated topics for all 14 topics. For each topic, I also include the Supreme Court opinion with the highest

---

7. Minimal preprocessing was used in order to prepare the texts for analysis. Specifically, I removed numbers and punctuation with the exception of intraword dashes, and I stripped capitalization from the corpus. In recent work by Denny and Spirling (2018), these choices were some of the least influential when it came to changing the substantive content of texts while vastly improving statistical inference.

8. I include only the most frequent, as—given Zipf’s law—words occurring at less than this frequency are generally quite rare and thus offer little help in identifying themes while slowing computation.

9. A number of algorithms are available for identifying the most representative terms for each topic. I use the FREX scoring algorithm, which identifies words that are both frequent within the topic and exclusive to that topic, relative to other topics.
### Table 1. Structural Topic Models and Their Characteristics

<table>
<thead>
<tr>
<th>Structural Topic Model</th>
<th>Case</th>
<th>Issue Area</th>
<th>Issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>writ_appeal_writof</td>
<td>Buckingham v. McLean, 54 U.S. (13 How.) 150 (1851)</td>
<td>Due process</td>
<td>Hearing or notice</td>
</tr>
<tr>
<td>tax_income_thetax</td>
<td>Kirby Petroleum v. Cmr. of Internal Revenue, 326 U.S. 599 (1945)</td>
<td>Federal taxation</td>
<td>Internal revenue code</td>
</tr>
<tr>
<td>indian_treaty_war</td>
<td>Roff v. Burney, 168 U.S. 218 (1897)</td>
<td>Civil rights</td>
<td>Indians</td>
</tr>
<tr>
<td>bankruptcy_creditors_suits</td>
<td>Russell v. Todd, 309 U.S. 280 (1939)</td>
<td>Federalism</td>
<td>Federal preemption</td>
</tr>
<tr>
<td>thestate_constitution_state</td>
<td>Mason v. Missouri, 179 U.S. 328 (1900)</td>
<td>Civil rights</td>
<td>Reapportionment</td>
</tr>
<tr>
<td>vessel_thevessel_ship</td>
<td>Kennedy's Claim (1814)</td>
<td>Due process</td>
<td>Takings</td>
</tr>
<tr>
<td>secretary_theact_thesecretary</td>
<td>McLean v. Fleming, 96 U.S. 245 (1877)</td>
<td>Economic activity</td>
<td>Trademark</td>
</tr>
<tr>
<td>employer_employees_labor</td>
<td>Malo Photo Supply v. NLRB, 321 U.S. 678 (1943)</td>
<td>Unions</td>
<td>Bargaining</td>
</tr>
</tbody>
</table>

Note.—Each entry indicates the topic generated from a 14-topic structural topic model and the case with the highest proportion for that topic. The second and third columns represent the broad issue area and the more specific issue, respectively, both as coded by the Supreme Court Database.
proportion of the opinion devoted to that topic. Finally, I include both the SCD’s assigned issue area code and issue code.

Note first the estimated topics generally reflect substantively interpretable themes based solely on the automatically generated three-word labels. For instance, “invention,” “patent,” and “machine” are the defining words for one topic. Other well-identified topics include a topic related to interstate commerce (“interstate, the commission, interstate commerce”), criminal activity (“criminal, conviction, crime”), and labor relations (“employer, employees, labor”). Moreover, the issue areas assigned by the SCD to these opinions consistently indicate that the automated approach identifies themes that closely match human understandings. In the case of the previous examples, the assigned issue areas are economic activity, economic activity, criminal procedure, and unions, respectively, while the more granular assigned issues are patent, regulation of railroads, death penalty, and bargaining, respectively. Importantly, higher dimensional topic models are capable of recovering these more granular representations for scholars interested in them.10

Three topics require additional discussion based on the assigned labels. Consider first the topic for which the topic label is “program, id at, at”; alternative approaches to identifying the top words for the topic identify “epa,” “erisa,” “afdc,” and “eeoc” as characteristic terms of the topic. That is, the topic directly relates to statutory language and the Court’s statutory interpretation of legislation. Likewise, for the “secretary, the act, the secretary” topic other principal terms include “postmaster,” “the treasury,” and “the collector,” indicating the association of the topic with judicial interpretation of executive actions. Finally, the “writ, appeal, writ of” topic is associated with other procedural legal actions, including “of error,” “depositions,” “the appeal,” “jurisdiction,” and “decree.” Taken together, three topics for which the assigned names prove initially difficult to interpret instead seem to reflect the underlying activities of the three branches of government and the continuing interpretive dialogue in which the Court is involved (see, e.g., Fisher 2004). In all, the estimates of the Court’s attention reflect sensible divisions of the Court’s attention, identified without imposing any structure other than a temporal component and providing proportional (rather than exclusive) estimates of attention.

The validity of the STM approach can also be seen in a plot of average attention to issues over time.11 In figure 1a, I plot the average attention to each SCD issue area, and in figure 1b, I plot the average attention to each of the 14 estimated topics. While sensible changes in attention according to the SCD are to be expected, the changes in attention according to the topic model are also sensible. Note, for instance, the near mirror similarity of patterns of attention as reflected by the SCD’s criminal procedure category and the estimated “criminal, conviction, crime” topic.

---

10. In the data release that accompanies this article available in the JLC Dataverse at https://doi.org/10.7910/DVN/YFLR9W, I include a 250-topic model, approximating the issue category of the SCD.

11. I use average attention per term in order to avoid falsely associating changes in the Court’s overall workload with changes in the Court’s proportion of attention devoted to an issue.
Figure 1. Changes in Supreme Court issue attention over time. A, Average issue attention by Supreme Court Database issue area codes. B, Fourteen-topic structural topic model.
Throughout, we see evidence that the estimated models generate topics sensible to human readers and coherently associated with concepts with which they should be associated (convergent validity). I turn then to a series of illustrations of the value added from such an approach for many research applications. As these illustrations demonstrate, while the prominent existing approach may at times be appropriate, for many other research applications the flexibility and specificity of topic models offer a more suitable measurement approach.

**ILLUSTRATION ONE: RELATED TOPICS**

One research application for which topic models and the associated topic proportions are especially valuable is when theory implicates or suggests relationships between topics. In those situations, the issue and issue area variables available from the SCD are, given single membership classification preferences, inappropriate. Instead, the simplification obscures important variations in behavior. Consider two recent articles. First, Harvey and Woodruff (2011) suggest systematic bias in issue coding introduced by a coder’s unconscious beliefs about the decision’s ideological direction. Here, confirmation bias manifests in decisions over which of many potential issue areas is primary. By manipulating the choice of the single issue classifications—as is required by the coding protocol—the coder shifts the ideological direction of a vote, thereby matching expectations of case disposition and justice ideology. Second, recent work by Rice (2016) suggests dissenting opinion authors strategically emphasize alternative issue dimensions on which the case could be decided. Although consistent with the sort of heresthetical maneuvering identified at all other stages of Court decision making (Epstein and Shvetsova 2002; Wedeking 2010; Black et al. 2013), the findings are contrary to other prior work suggesting the Court primarily divides (and decides) solely on a single issue dimension in the overwhelming majority of cases. In both articles, the researchers document how the coding protocol—and specifically the unidimensional assumption—of the dominant measure of opinion content and Supreme Court issue attention yields measures of issue attention that are inappropriate for the purposes of some empirical research settings.

Alternatively, topic models and the associated topic proportions enable researchers to directly estimate the relationships between topics and to examine variations in topic attention within opinions. Consider the relatively simple case of correlated topics. As an example, in *National Federation of Independent Business v. Sebelius* (567 U.S. 519 [2012]), the Court considered—in addition to jurisdictional questions—three primary issues: (1) the constitutionality of the individual mandate under Congress’ commerce clause powers, (2) the constitutionality of the individual mandate under Congress’ taxing and spending clause powers, and (3) the constitutionality under principles of federalism of the withholding of federal Medicaid funding to states that do not establish exchanges. Importantly, these legal issues are frequently linked to one another in cases, as claims of federal infringements on state power often follow from congressional exercises of power under the commerce clause or taxing and spending clause. Indeed, leading textbooks for constitutional law often explicitly join the issues together (e.g., Epstein and Walker 2017).
Identifying such correlated topics—well understood by legal scholars and social scientists but hard to identify when preferential treatment is given to assigning a single, mutually exclusive topic—is simple with STMs, which permit positive correlations among the estimated topics. To see this, in figure 2, I plot a heat map of the correlation matrix among topics. Each entry in the heat map reflects the correlation coefficient between the identified row and column topics. In the plot, light gray indicates negatively correlated topics, while dark gray indicates positively correlated topics, with the size of the square proportional to the magnitude of the correlation. The topics are ordered according to a hierarchical clustering of the resulting correlations, with black boxes indicating the identified clusters. Before proceeding, note that if the same plot were drawn with the single mem-

12. Clusters are estimated using single-linkage agglomerative clustering.
bership treatment of the SCD’s issue areas, each of the off-diagonal elements would be white, as they are effectively treated as zero.

Instead, with the topics estimated from an STM we can identify a series of interesting relationships in the themes characterizing the Court’s opinions. First, each of the clusters offers evidence of a coherent area of policy attention on the Court. The three main clusters feature topics that are strongly linked together. Notably, the clusters identify a group of correlated topics related to civil rights and liberties and another cluster related to economic issues. Yet although the topics cluster, one can still see positive correlations outside of the clusters as well; note, for instance, that the interstate commerce topic (“interstate,” “the commission,” “interstate commerce”) is positively correlated with a topic related to congressional power (“use,” “cong,” “sess”) as well as a topic related to federalism and state power (“the state,” “state,” “constitution”). Turning back to the example of NFIB v. Sebelius, one can see that across the corpus of opinions—the history of Supreme Court attention—the topics from which the Court’s questions arose consistently and positively relate to one another. In other words, discussion of one of these topics frequently accompanies the discussion of another.

Of course, increasing the number of topics would permit identifying even more specific attention relationships in Supreme Court opinions. Even at this general level, though, finding that topic proportions for theoretically connected concepts correlate across the corpus provides yet further evidence of the value in considering opinions as composed of topic proportions—rather than a single issue—for some research endeavors.

ILLUSTRATION TWO: CASE COMPLEXITY

Topic proportions also offer a robust method for addressing case complexity, a concept regularly cited in law and politics research (e.g., Maltzmann, Spriggs, and Wahlbeck 2000; Collins 2008; Carrubba and Zorn 2010) but with little resulting consensus (e.g., Goelzhauser and Vouvalis 2014). Part of this arises from the fuzzy nature of the concept, with little effort in prior research to clearly explicate a clean definition of the concept of case complexity. Breaking from this general pattern, Goelzhauser and Vouvalis (2014, 3) define complexity as referring to “the amount of information required to process before reaching a decision.” This definition neatly characterizes most of the treatments of the concept in prior work.

As an example, I turn to Lax and Rader (2015), in which the authors model the influence of opinion authors over the maintenance (or defection from) the majority coalition. The authors hypothesize that there will be more defection from the majority coalition when cases are more complex. The underlying theory suggests that cases with more areas for disagreement generate more difficult bargaining conditions and thus increase the probability of defection. In order to capture case complexity, they use a dichotomized version of the SCD’s coding of legal provisions, with 0 indicating a case with only one legal provision and 1 indicating a case with more than one legal provision.
Alternatively, consider topic models and the associated topic proportions in light of the theoretical underpinnings of case complexity. Definitions of case complexity focus on connecting the concept to the amount of information that needs to be processed. With topic proportions, the amount of information processed is laid bare. Each opinion is divided into the proportion of attention devoted to a particular topic; when attention is concentrated within a single topic area, little other information has been processed by the justices. But when attention is diffuse across multiple topics, the Court is directly signaling the additional information that was necessarily gathered, considered, and evaluated in reaching the decision.¹³

Estimating concentration or diffusion of proportions is a well-studied area in economics and other related fields, developed to study, for instance, market shares. A standard approach to estimating concentration is the Herfindahl-Hirschman index (H-index), which reflects the extent to which attention is devoted to a single or multiple topics and is equal to

\[ C_d^* = \frac{\left( C_d - \frac{1}{N} \right)}{1 - \frac{1}{N}}, \]

where \( C_d = \sum_{i=1}^{N} \hat{p}_i^2 \) is the standard H-index, \( N \) is the number of topics, and \( \hat{p}_i \) is the proportion of a topic captured in a given document. In order to enhance comparability, I multiply the topic concentration by \(-1\); thus, higher values indicate documents that address many different topics, whereas lower values indicate documents more concentrated on a single topic.

I construct measures of topic concentration across 15-, 25-, and 50-topic STMs. With the measure of complexity based on topic concentration in hand, I replicate Lax and Rader (2015). In each of the analyses, the dependent variable is whether or not the member of the majority coalition defected from the majority. I retain all covariates from Lax and Rader (2015) in the analysis. Importantly, for every variable other than the case complexity measure, all substantive results persist across specifications.

However, in not one of the three replications does case complexity reach standard levels of statistical significance. To demonstrate the difference across measures of case complexity, I estimate predicted probabilities across the range of values for the relevant measures—a dichotomy of the number of legal provisions in the original article and an alternative measure based on topic concentration. The results appear in figure 3. Because the results across topic dispersion–based measures of complexity are consistent across the number of topics, I present only the results from the 15-topic specification. In the left panel of figure 3, the increase in the probability of defection as one goes from one legal provision to more than one legal provision is clear. But the change in the predicted probability across the entire range of values for topic dispersion makes clear the lack of a statistically or sub-
stantively significant effect from case complexity on the probability of defection. Thus, with a more granular, continuous measure of case complexity, one finds little evidence that case complexity influences the probability of defection.

**ILLUSTRATION THREE: ROBUSTNESS**

Finally, the flexibility of topic model estimation offers an avenue by which to assess the robustness of empirical results across alternative specifications. The replication crisis in scientific research is well documented (Gerber et al. 2010; Simmons, Nelson, and Simmons 2014). Here, identifying a single topic structure potentially obscures important variation within a topic. Although there are often theoretical reasons for preferring smaller or larger numbers of topics within an individual research project, present research is frequently resigned to employ the SCD’s issue area codes. Yet it is well understood that the choice of the number of topics may influence inference; indeed, this is a subject of great debate in the topic modeling community (Wallach, Mimno, and McCallum 2009). Because of the ease and flexibility of estimating models, however, estimating models across values of the number of topics offers an additional robustness check for analyses.14

To demonstrate, I turn to research on the litigant signal model (Baird 2004, 2007; Baird and Jacobi 2009a, 2009b), which suggests the Supreme Court—in deciding cases—signals litigants as to areas it considers a priority. In responding to these signals, litigants substantively shift the pool of available cases available to the Court for review, providing the Court with an additional agenda-setting influence and yielding a more powerful policy-making institution. The research identifies significantly more court cases 2 years

---

14. For an example of this in an applied setting, see Rice (2016).
and even more prominently 5 years after the increase in the Supreme Court’s prioritization of an issue area.

The dependent variable in Baird’s 2004 analysis is the number of cases decided by the Supreme Court during a term. The independent variable of interest is a measure of policy priority, which Baird captures by aggregating the number of salient cases within a term for each policy area. Salience is measured through the (Epstein and Segal 2000) measure of case salience, which is simply an indicator for whether or not the decision of the Court was featured on the front page of the New York Times the day after the decision.

Thus, both the dependent and primary independent variables in the data set are counts of cases within a given policy area during a year. In that context, defining the boundaries of policy areas is obviously critical. Baird identifies 11 issue areas, relying almost entirely on the classification scheme of the SCD but making some minor modifications (e.g., combining some categories while dropping others from the analysis). Because the exact time period necessary for the Court’s agenda to respond to the additional litigant attention is unclear, Baird specifies an autoregressive distributed lag model, with six lags of the primary independent variable, controls for Rehnquist and Burger Court eras, and a lagged dependent variable to account for temporal dynamics. In so doing, the specification of the model features six independent variables all directly tied to the identification of policy, in addition to a series of fixed effects for each policy area.

To again demonstrate the utility and flexibility of STMs, I turn to replicating the above analysis but with topic attention instead as estimated from a series of STMs. Specifically, I estimate six STMs, with \( k \) set to values of 5, 10, 15, 20, 25, and 50. Rather than case counts for each policy area, I sum the proportions of attention devoted to the policy area for all cases (the dependent variable) and for salient cases (the independent variable), with salience again captured through the New York Times measure.\(^{15}\) Thus, the measure more accurately captures the Court’s total attention to a policy area, as subordinate attention is explicitly incorporated.

In figure 4, I plot a visualization of the relevant model results. Specifically, for each model specification, I plot the coefficient estimate for each of the six lags of political priority and the associated 95% confidence interval. Two observations bear discussion. First, the general result persists across a series of specifications; that is, throughout the models we see that increases in the saliency of the Court’s attention to a policy correspond with increases in subsequent Court attention to that policy.

Second, however, the lag at which we would conclude there is a relationship varies across every estimation. The most consistent lags are the 3- and 5-year lags, both of which are positive and significantly related to subsequent attention in all but one specification. That inconsistency makes it difficult to conclusively settle on the particular

---

\(^{15}\) Consistent with Baird, the models include fixed effects for policy area. Rather than chief justice fixed effects, I included fixed effects for the October term, which I was able to do here because I did not face the model identification problem confronting Baird’s specification.
Figure 4. Coefficient estimates (black points) and associated 95% confidence intervals (gray bars) for the effect of lags of Supreme Court issue prioritization on future Supreme Court issue attention, using topic estimates. Each plot is estimated based on topic attention measures from a structural topic model with the indicated number of topics.
length of time at which the Court’s prioritization of a policy might generate subsequent. Indeed, at larger numbers of topics nearly every subsequent year approaches or exceeds standard levels of statistical significance, suggesting the Court’s agenda-setting influence is important but temporally dispersed. As before, substantive inferences are adjusted based on incorporating the additional information available from the mixed membership classification and doing so across a variety of specifications.

**DISCUSSION AND CONCLUSION**

In this article, I introduce a series of measures based on a now well-accepted approach from computational linguistics for capturing textual content. For many research applications, these proportions offer marked improvements over existing measures. First, they overcome data lost through single membership classification by instead providing estimates of the proportion of each document addressing each topic. In so doing, the approach better reflects opinion content in ways consistent with the movement of the field toward the strategic model and more complex assessments of judicial behavior. Moreover, they offer a flexible approach for scholars interested in studying the Court’s attention, as they can be simply adjusted and estimated in order to yield more general topical frameworks or highly specific frameworks. To that end, all texts and code necessary for scholars to replicate and extend the models for their own purposes are publicly available, enabling future research to incorporate more sophisticated understandings of opinion content when necessary to improve causal inference. Moreover, for the researcher who wishes to use these topics to classify new opinions into identical topics, the estimated models are available, providing an avenue to directly compare results across updated corpora.

Although the above demonstrates the value of topic models, if anything I understate the potential of topic models, and in particular STMs, for empirical legal research. As noted previously, STMs can include covariates that help to structure and define either topical prevalence, or how prevalent the topic is within a document, and topical content, or what words are used within a topic. This effectively allows one to analyze the influence of different covariates of interest on the topic distributions. In the context of the above models, I have estimated the effect of time on the prevalence of topics, which helps to address temporal dynamics. But consider alternative specifications, for instance a variable specifying the Rehnquist Court. Here, one can then examine the effect of those covariates on topical prevalence or content. In the case of a Rehnquist Court variable, one might expect to see a specific effect on federalism. In other areas, recent work has leveraged similar approaches in analyzing differences in open-ended survey responses (Roberts et al. 2014). For the Court, the opportunities are myriad. For instance, one could even extend the Rehnquist Court example, potentially analyzing differences precipitated by ideology on the prevalence of federalism arguments or on the language associated with federalism arguments.

Finally, moving forward, research might look to build on the approach identified here to construct increasingly sophisticated indicators of the content of legal documents. The
approach used here should be seen as an important first step in this process, but the effort to identify and measure opinion content should certainly continue. Legal texts are—as legal scholars have frequently reminded social scientists (Friedman 2006)—much more complex than often presumed in empirical research. Alternative methods for estimating topics from text corpora and minor improvements and adjustments to existing approaches continue to be rapidly introduced across a variety of technical and applied fields of research, and as these tools become established it may become clear they offer yet further improvements on that which I have introduced in this article. I leave this to future research.

REFERENCES


QUERIES TO THE AUTHOR

Q1. Your article has been lightly copyedited for journal style, consistency, grammar, and clarity. Please read your page proofs carefully to make sure we did not inadvertently alter your meaning. Note that we may be unable to make changes that conflict with journal style or cause grammatical problems.

Q2. Footnote 2: Please confirm that the URL is correct. It is the same as the one for the Dataverse mentioned in the acknowledgment footnote.

Q3. Footnote 3: Please provide a URL where the public discussion can be found.

Q4. “classifies instead legal issues”: I removed the italics from “legal,” as journal style discourages the use of italics (or any other device) for emphasis. You may wish to restore your italics here and there, but keep in mind that less is more in this case.

Q5. “as the coding protocol states”: Please add to the reference list a citation to the coding protocol.

Q6. “the public availability of”: Delete “the public availability of” as redundant? Otherwise, you say “the availability of . . . is available.”

Q7. “thus all subordinate issues—although potentially”: Is a word missing after “potentially”?

Q8. “economic activity, economic activity”: Is the repetition here intended?

Q9. Footnote 11: Please confirm that the URL is correct. It is the same as the URL for the Dataverse given in the acknowledgment footnote.

Q10. “might generate subsequent”: Is a word missing after “subsequent”?

Q11. Roberts et al. 2013: Could you provide a bit more information about the Proceedings? Do they exist online? Have they been published in book form and hence is there a city of publication and name of publisher to list?

Q12. Hopper v. Evans 1981: Please confirm that the date was 1981. I saw 1982 online.


Q14. Russell v. Todd 1939: Please confirm the date. I saw 1940 online.


Q16. Kennedy’s Claim: Can you cite the United States Report number, as in the other cases?

Q17. Howard County 1882: Please confirm the date. I saw 1883 online.

Q18. Medo Photo Supply 1943: Please confirm the date. I saw 1944 online.