Power in Text: Extracting Institutional Relationships from Natural Language

Robert Shaffer

May 1, 2017

Abstract

How do legislators allocate policy-making authority? Generally speaking, institutional design decisions involve a trade-off between efficiency and accountability, as legislators seek to simultaneously maximize bureaucratic effectiveness and ensure favorable policy outcomes. At least in the legal context, these design decisions are often articulated in textual documents (e.g. statutes and constitutions). Unfortunately, existing measurement schemes cannot capture the full range of institutional design technologies available to legislative actors. These limitations have prevented scholars from addressing important questions regarding the relationship between executive/legislative preference conflicts, background institutional context, and downstream design of legislation. In this paper, I develop a text-based measurement scheme intended to address these limitations, which I apply to an original dataset of American legislative texts.

1 Introduction

Why do lawmakers write complex laws? As any would-be lawyer knows, legal documents (e.g. contracts, statutes, or constitutions) are famously difficult to read, and more difficult still to write. Much of this difficulty is *structural* in nature; in large part, legal texts are designed to structure relationships between actors, describing who can do what, when, and to whom. In some cases, the structures envisioned by a particular text are straightforward, but they can also be remarkably complex. To take a pair of extreme examples from American politics, the Clean Air Act grants the Environmental Protection Agency sole decision-making authority over most policy decisions, while the Affordable Care Act divides decision-making authority between the Departments of Labor, Treasury, Health and Human Services, and Veterans' Affairs, as well as various subordinate agencies and other institutions.

As numerous authors (e.g. Moe 1990b; McCubbins 1985) have noted, designing complex decision-making structures is costly, both for legislators and for implementing bureaucrats. As a result, we might wonder why (at least in some situations) lawmakers choose to pay these costs. Particularly in the American context, an array of theoretical and applied researchers have studied this question, and have largely focused on versions of the "ally principle" as the key explanatory variable (Epstein and O'Halloran 1999; Huber and Shipan 2002; Bendor and Meirowitz 2004; Franchino 2007; Gailmard and Patty 2012; Farhang and Yaver 2016). When the interests of the executive branch are aligned with those of the legislature (and are likely to remain so in the future), these authors argue, legislators are more willing to pass simple, "framework" legislation, and leave implementation choices to the bureaucracy. By contrast, when legislative and executive interests diverge, legislators tend to favor complex institutional structures, which provide substantial oversight opportunities and divide decision-making authority among an array of implementing bodies.

All of these studies are important, and all have enhanced our understanding of the process of statutory design and implementation. All, however, also share significant limitations. As I argue in this paper, formal (text-based) allocation of authority is best viewed as a *relational* process, in which law-writers seeking to implement a policy program simultaneously select a group of implementing actors and a set of decision rules governing the relationships between these actors. Both of these choices are informed by background political and policy-related factors (e.g. preference disagreements between executive and legislative actors), and both have important consequences for downstream policy outcomes. Unfortunately, measurement and conceptual limitations have prevented existing studies from capturing the full range of institutional design decisions involved in formal allocation of authority. Instead, most studies focus on imprecise conceptual ideas (e.g. the overall "level of delegation" offered by a particular statute) which obscure the causal relationships and outcomes of interest.

In this paper, I introduce a novel conceptual and measurement approach designed to address these shortcomings. Since many decisions regarding distribution of authority and regulatory attention to particular policy areas are explicitly articulated in text documents (e.g. legislation, administrative rules, constitutions), language-based information extraction offers an attractive approach. After reviewing the existing state of the literature, I develop and present a grammar-based institutional parser, which allows users to identify and extract institution mentions and relationships from legal text. This technique effectively treats each document as a network of inter-actor relationships, which allows researchers to answer questions regarding the relationship between background political and preferential factors and downstream allocation of authority. Finally, I apply this approach to a novel dataset of US legal documents, and discuss directions for future work.

2 Policy Design and Text Analysis

2.1 The Existing Work

In the political science and legal literatures, formal (textually-defined) distribution of decision-making authority is a key phenomenon of interest. Frequently, social scientists are primarily interested in studying the inputs, outputs, and performance of some organization of interest (e.g. economic performance of a business, or policy choices by a government). Clearly, decision-making structures play a critical role in research programs of this sort. As Kagan (2009) argues, for example, compared with those of other countries American laws are unusually complex and restrictive, splitting authority between a broader set of actors and encouraging a more adversarial mode of dispute resolution (see also, e.g., Farhang 2010; Kelemen 2011). In policy terms, the impact of this choice of regulatory mode is not obvious; at the very least, however, the decision to create complex, "fragmented" authority structures involves a trade-off between policymaking efficiency and regulatory accountability (Farhang and Yaver 2016).

Generally speaking, scholarly work on the design of formal institutions has treated the phenomenon as a principal/agent problem, in which lawmakers are forced to balance efficiency and responsiveness against oversight and popular responsiveness. As Moe (1990a; 1990b) and Moe and Caldwell (1994) note, high levels of detail and policy complexity are usually not desirable. Detailed prescriptions, complicated power structures, and onerous procedural requirements all dramatically curtail administrative flexibility, reducing implementer responsiveness and promoting policy gridlock. Moreover, policy restrictions are not costless to create. If policymakers want to insulate their programs from executive malfeasance, they need to design the relevant restrictions carefully, listening to expert evidence and establishing the appropriate incentive structures. In complex policy areas, where the links between policy and outcomes are more uncertain, these problems are particularly severe (Bendor and Meirowitz 2004; Epstein and O'Halloran 1999, 73-75, 84-85). Using "fire alarms" (McCubbins and Schwartz 1984) to delegate oversight and protection functions to interest groups may help somewhat, but fire alarms and other ex post restrictions are not always effective (Epstein and O'Halloran 1999; McCubbins et al. 1989). Thus, at least in the abstract, policymakers prefer streamlined, simple laws, which offload most decision-making authority.

Here, though, Moe's "politics of structural choice" intervenes. Whenever principal (legislature) and agent (executive) disagree on substantive preferences, insulation becomes more attractive as a policy strategy, protecting relevant programs from implementer malfeasance (McCubbins 1985). Moreover, as uncertainty about future electoral outcomes rises, restrictive policies become particularly appealing; if a currently-empowered group is not sure if it will hold public office in the future, that group will likely try to passively "insulate" its programs, delegating less authority than it otherwise might (Ibid; Moe and Caldwell 1994; de Figueiredo Jr 2002). Insulation is thus particularly appealing for electorally weak groups and members of public interest organizations, who are likely to hold public office only sporadically.

Based on these ideas, scholars have investigated an array of potential influences on institutional design decisions. Generally speaking, these scholars have used one of two measurement approaches; specifically, hand-coding schemes (e.g. Epstein and O'Halloran (1999) "discretion index" or Farhang and Yaver (2016)'s "fragmentation index"), and automation-friendly heuristics (e.g. Huber and Shipan's (2002) word-count measure). Studies using hand-coding schemes, for their part, usually rely on third-party legislative summaries to extract relevant information, which use to form a measure of some quantity of interest. For example, Epstein and O'Halloran (1999) use Congressional Quarterly's year-end legislative summaries to count the number of "major provisions" (scope) contained in their statutes of interest. They then count the number of constraints placed on the exercise of those powers, and use those two values to calculate each statute's "discretion index" (Epstein and O'Halloran 1999, 100-108). Franchino (2007) and Farhang and Yaver (2016) use a similar approach to measure "delegation" and "fragmentation", respectively.

Other studies, by contrast, adopt more automation-friendly approaches. Most prominently, Huber and Shipan (2002) argue that "when faced with two statutes that address the same issue, the longer one typically places greater limits on the actions of other actors" (Ibid 2002, 45). Using this observation, they operationalize discretion using document word counts, and examine a broad set of statutes related to American state-level Medicaid expansion and European labor regulation.

Different scholars have used these methods across a variety of contexts, with some success. Various writers have found that legislative capacity, the availability of informal restrictions on delegation, and issue complexity are all significantly related to delegation levels (Epstein and O'Halloran 1999; Franchino 2007; Huber and Shipan 2002). Similarly, many of these same studies find that legislatures operating under divided and minority government delegate less authority than in situations of unified institutional control (e.g. Oosterwaal et al. 2012). Legislators respond similarly to expected changes in the electoral environment (Ainsworth and Harward 2009). For the most part, these studies have used parties as the unit of analysis; however, at least in the US context, there is some evidence that the majority party's seat share matters as well (Epstein and O'Halloran 1999, 136).

2.2 Gaps and Limitations

Based on the existing scholarship, we now possess strong knowledge about patterns in formal allocation of authority across a variety of institutional contexts. However, despite these advances, significant gaps remain. Most prominently, measurement limitations have prevented authors from examining the relationship between characteristics of policy area and downstream institutional design. Take distribution of lawmaker attention; as Baumgartner et al. (2009, 608) note in their discussion of attention dynamics,

> As a given social indicator becomes more troubling over time, the bounded rationality model predicts no response whatsoever during the early periods; the issue is "under the radar"' and government may not even track its severity in any systematic manner. After the severity of the issue has passed some threshold, on the other hand, there may be a rush to make up for past inattention to the issue by dramatically increasing policy outputs directed to it. The issue may be systematically tracked and a specialized agency or bureau may even be created to focus on it.

Institutional design decisions likely follow a similar pattern to other government policymaking decisions. Since complex institutional systems are costly to design, laws that address low-attention issues are likely to create straightforward systems that leave most decision-making powers to the relevant administrative agencies.

Unfortunately, because of measurement limitations, previous studies in this area have been largely unable to test this prediction. Hand-coding methods, for their part, require a substantial time and labor investment, making comparisons across many different policy areas and impractical. Word counts and similar document-level measures are easier to scale, but do not generalize easily to inter-group comparisons. Even within a particular legal system, laws addressing one policy area may be more verbose than those in another policy area, regardless of their content. In order to test these hypotheses more thoroughly, then, we need a different strategy.

2.3 Towards a Relational Conception of Formal Power

To borrow from Dahl (1957), whenever A provides authority to B, A faces at least two choices. First, what powers should A allow B to exercise? And second, what restrictions should A place on B's decision-making powers? Most decisions involving formal allocation of authority can be viewed as a combination of these two decision types. For concreteness, consider the following examples:

1. The Secretary [of the Interior] shall [...] determine whether any species is an endangered species or a threatened species because of any of the following factors:
   (a) the present or threatened destruction, modification, or curtailment of its habitat or range;
   (b) overutilization for commercial, recreational, scientific, or educational purposes[1]
2. The Administrator for Federal Procurement Policy, in consultation with the Secretary of Defense, the Administrator of General Services, and the Director of the Office of Personnel Management, shall develop and implement a plan to ensure that the Federal Government maintains the necessary capability with respect to the acquisition of architectural and engineering services.[2]
3. After a Money Bill has been passed by the House of the People it shall be transmitted to the Council of States for its recommendations and the Council of States shall within a period of fourteen days from the date of its receipt of the Bill return the Bill to the House of the People with its recommendations and the House of the People may thereupon either accept or reject all or any of the recommendations of the Council of States.[3]

These three provisions come from different countries (the United States and India) and different types of legal documents (one national constitution, and and two ordinary statutes). All three, however, offer good examples of the kinds of language common in documents that formally allocate authority. Like all legal documents, each of these texts addresses a certain set of topics and policy areas, empowering certain actors to use particular kinds of instruments and address particular issues (McCubbins 1985, 726). Each text also imposes a certain set of constraints upon the actors it mentions. The first excerpt, for example, provides simple decision-making authority to the Secretary of the Interior, with policy-based directives regarding the types of evidence to be considered as part of that decision. The second excerpt is collaborative, replacing outcome-based policy directives with an overlapping institutional structure designed to represent the various agencies and groups involved in defense procurement policy. Finally, the third excerpt mixes several constraint types, providing both a procedural requirement - namely, a deadline - as well as a set of relational constraints on the procedures for enacting money bills.

All of the language contained in these excerpts is important, and provides information about the relationship between political factors and downstream policy implementation. However, for the purposes of this paper, I focus on *relational* language, or language in which the limits on powers exercised by particular actors are determined by the formal institutional relationships defined in the text. Legal texts, from this

[1] 16 U.S.C. 1533(a)(1-2)
[2] 41 U.S.C. 1128
[3] Constitution of India (1950), Art. 109(2).

perspective, can be viewed as a network of institutional institutional relationships whose characteristics (e.g. number of nodes; density of ties; presence or absence of communities) can be manipulated by legislators in order to advance particular goals. McCubbins (1985, 725-726) provides a useful summary of types of relational constraints present in bureaucratic policymaking:

> Regulations can be administered through civil or criminal suits in the courts, or through independent commissions or executives agencies, through discretion granted to the president or state and local entities [...] The choice of institutional setting by [a legislature] involves a decision on how much independence [that legislature] wishes to grant the administrators (independence from [legislative] control) and the extent to which other decision-makers [...] restrict or influence the choices of the administrators.

Though less concrete than outcome-based directives or procedural requirements, these network-based institutional restrictions can be remarkably powerful. By manipulating the background institutional structure, legislators can force implementing actors to cooperate with other (often hostile) players, altering downstream policy outcomes and restricting implementer discretion.

To be clear, this focus on the institutional relationships articulated by legal texts is unique to this paper. For example, Epstein and O'Halloran (1999)'s "discretion index" and Farhang and Yaver (2016)'s "fragmentation index" can both be viewed as properties of the implicit institutional networks established by the texts these authors examine. However, by placing network characteristics at the center of our conceptual scheme, we can generate natural measures for key ideas like policy complexity, clarifying the connection between data and underlying quantities of interest.

3 Information Extraction

The relational conceptualization of formal power I give in the previous section leads to a natural set of research questions. As noted earlier, creating complex, institutionally interconnected statutes is costly, requiring politicians to expend time and energy consulting with experts and drafting intricate legal language. Under what circumstances should we expect legislators to be willing to pay these costs? Previous research in the literature on formal institutional design suggests that legislators are most willing to delegate authority when their preferences diverge from the executive's; so, legislators ought to be most willing to pay drafting costs under divided rather than unified government. The relationship between policy area and legislative complexity is less clear, but at the very least, in higher-salience and higher-attention policy areas we ought to expect legislators to be more willing to pay appropriate drafting costs.

In order to explore these questions further, however, we need an appropriate measurement scheme. As I argue in this section, natural language processing (NLP) provides a promising general approach. NLP is a catch-all term for a set of computational methods that attempt to analyze the linguistic attributes of a given text. NLP methods are thus extremely wide-ranging, covering topics like part-of-speech tagging, grammatical parsing, lexical co-occurrence, latent content analysis (e.g. topic modeling), and much else besides. Coupled with a deep, politically- and legally-informed understanding of the documents in question, these tools can provide a powerful approach to the analysis of legal texts.

3.1 A Grammatical Information Extraction Approach

Since the key conceptual idea I emphasize in this paper is a *relational* notion of formal power, the measurement scheme associated with this conceptualization needs to be able to extract the appropriate institutional information from the text. This measurement process can be summarized in two basic steps:

1. **Entity recognition.** Mentions of relevant actors (e.g. "President" or "Secretary of the Interior") need to be identified and recorded, along with any contextual and identifying information.
2. **Relation extraction.** Relationships between actors need to be defined and extracted. These relationships may be simple (e.g. co-occurrence in predefined units of text) or complex (e.g. hierarchical relationships between actors).

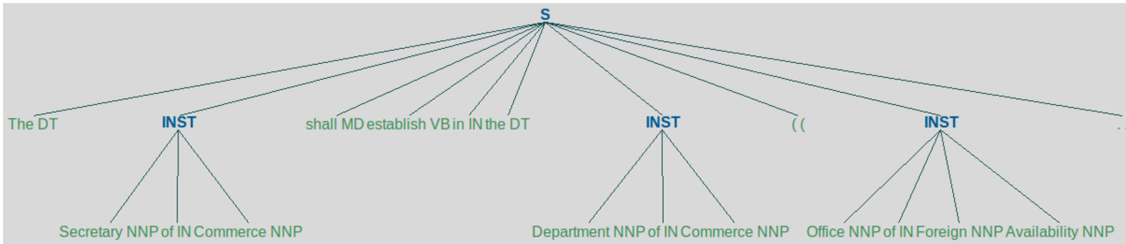
Both of these measurement steps have a rich history in the NLP literature, with an extensive history and available set of methods. For the purposes of this paper, I focus on using grammatical information as an information extraction technique. As a motivating example, consider the following sentence:

> The Secretary of Commerce shall establish in the Department of Commerce an Office of Foreign Availability, which shall be under the direction of the Under Secretary of Commerce for Export Administration. The Office shall be responsible for gathering and analyzing all the necessary information in order for the Secretary to make determinations of foreign availability under this chapter.[4]

These sentences lay out a fairly complex administrative structure. The text directs the Secretary of Commerce to create a new administrative agency within the Department of Commerce regulating export of defense technology, which is answerable to an Under Secretary. To capture this structure, we therefore need to recognize and record the four entities contained in this text, and match their shortened forms to their full forms (e.g. "Office" vs "Office of Foreign Availability").

[4] 50 U.S.C. 4604(f)(6)

Figure 1: Sample Parsetree



Partial parsetree for example language given above. “INST” refers to to term groupings identified as named entities by the parsing algorithm described in text. Other labels refer to part-of-speech labels assigned by the part-of-speech tagger.

Table 1: Sample Entity Extraction

Sentence	Raw Entities	Processed Entities
1	Secretary of Commerce	Secretary of Commerce
1	Department of Commerce	Department of Commerce
1	Office of Foreign Availability	Office of Foreign Availability
1	Secretary of Commerce for Export Administration	Secretary of Commerce for Export Administration
2	Office	Office of Foreign Availability
2	Secretary	Secretary of Commerce for Export Administration

Entities extracted by the algorithm described in-text, pre- and post-processing.

3.1.1 Entity Extraction

Entity extraction is a classic NLP problem, and many approaches are possible. Here, however, I use a grammar- and heuristic-driven approach. First, I use a pre-trained part-of-speech tagger[5] to apply part-of-speech labels (e.g. “Proper Noun (NNP)) to each term in the text. Then, I define a custom proper noun grammar (given below), and use a regular expression parser to extract phrases matching this grammar from the text. I then filter these phrases using white- and black-lists of inclusionary terms (e.g. “Secretary”, “Department”) and exclusionary terms (e.g. “Act”, “Law”).[6] Finally, to match shortened names to full terms, I adopt a heuristic approach, in which I assign single-word whitelisted proper nouns to the most recent extended version present in the text (e.g. “Secretary” to “Secretary of Commerce”).

[5]As implemented in Python’s NLTK library, available at http://www.nltk.org/.
[6]See Appendix for details.

The results of this algorithm as applied to the motivating example above are given in Figure 1 and Table 1. In this case, a rules-based approach performs reasonably well. All named entities given in the text are identified, and all terms associated with each term are extracted accurately (with the exception of the word "Under" in "Under Secretary of Commerce"). However, the heuristic I use to match the shortened term "Secretary" to its full definition ("Secretary of Commerce") fails in this case, matching "Under Secretary" instead. In future work, I plan to use standardized definition sections more explicitly in these cases. I also well as to expand the pre-defined white- and black-list terms used to filter incorrect matches to include "similar" terms to those given in the pre-defined list I use (e.g. using embedding vectors generated by word2vec (Mikolov et al. 2013)). For now, though, this approach provides a reasonably effective starting point.

3.1.2 Relation Extraction

Compared with entity extraction, relation extraction is a substantially more complex problem. Given a predefined set of training examples for a particular relation type, good machine-learning approaches to generalize relational labels exist (e.g. Roth and tau Yih 2004; Surdeanu et al. 2011). However, as I describe in the conceptualization section of this paper, a key variable of interest in studies of formal allocation of authority is the *density* of inter-actor relationships, rather than the nature of those relationships. Conceivably, we might be able to articulate hypotheses and theoretical expectations regarding the prevalence of particular inter-actor relationships (e.g. hierarchical or collaborative structures), but these ideas are less intuitive, and have fewer connections with the existing literature. For now, then, I focus simply on the complexity of the documents under consideration (as measured by the prevalence of inter-actor ties) rather than modeling particular relationship types.

As in the previous section, I take a relatively simple heuristic approach to extract the relationships of interest. As noted in the drafting guide for the US Consolidated Code, the "basic unit" of every section of Code and legislation is the Section.[7] Laws and Consolidated Code fragments can be subdivided arbitrarily, but sections are intended to be stand-alone units that are roughly comparable in substantive scope. As a result, if we observe that two actors are co-mentioned in a Section of a law, we can reasonably conclude that those two actors share authority over the policy area under consideration in that section. Without a sharper definition of the relationships under consideration, we cannot draw strong conclusions about the nature of the connections between these actors, but we can draw general conclusions about the complexity of the structures created by a given law.

The relation extraction procedure I employ in this paper, then, proceeds in three basic steps. First, I segment each text according to its internal organization (e.g. titles, sections, etc), and remove preliminary material, section titles, and section headers

[7] http://uscode.house.gov/detailed_guide.xhtml

from the text.[8] Next, I remove amending text from each document. Amendments to existing law are not likely to follow the same drafting rules as ordinary legal text and often break formatting rules or cover a more diverse set of policy areas, leading to comparability issues. Finally, I recombine each text into sections, extract entities from each section, and draw an edge between any set of entities that co-occur in a given section. Using these data, I can then calculate a modified measure of network density:

$$D_i = \frac{E_i}{\sum_{j=1}^{N_i} \sum_{m>i}^{N_i} min(n_j, n_m)}$$

Where E_i is the number of ties actually present in document i , N_i is the number of unique entities in document i, and n_j is the number of observations of entity $j \forall j \in \{1, 2, ...N_i\}$. This formula simply represents the number of observed edges in the data, divided by the number of potential edges (excluding cases in which an entity is connected to itself). For a concrete example, consider the following (hypothetical) data:

1. The **President** shall nominate **Supreme Court Judges**.
2. **Congress** shall approve the **President**'s nominees.

Named entities are bolded. In this case, there are three unique named entities ("President", "Congress", and "Supreme Court Judges"), with "President" mentioned twice and both other choices mentioned once. Two edges are observed ("President/Supreme Court Judges" and "Congress/President"), and three are possible ("Congress/Supreme Court Judges" is not observed). This gives us the following density value:

$$\begin{aligned} D_i =& \frac{2}{\sum_{j=1}^{3} \sum_{m>j}^{3} min(n_j, n_m)} \\ =& \frac{2}{1+1+1} \\ =& \frac{2}{3} \end{aligned}$$

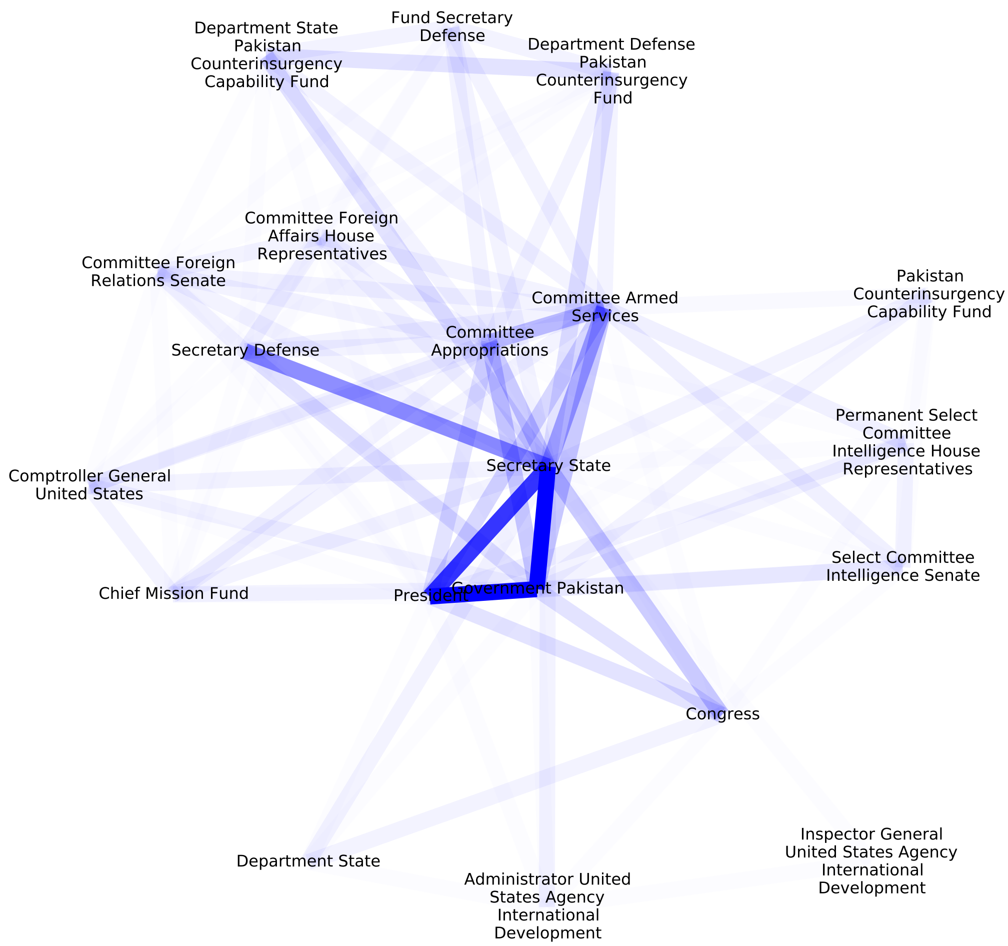
4 The Design of American Law

4.1 Data Collection

As a test of this approach, I collected an original dataset of named entity and relation data, consisting of all legislation enacted by the American Congress from

[8] Using the parser contained in https://github.com/rbshaffer/constitute_tools

Figure 2: Enhanced Partnership with Pakistan Act of 2009



Sample network graph for the Enhanced Partnership with Pakistan Act of 2009. Edges are shaded based on the nmber of ties between connected nodes.

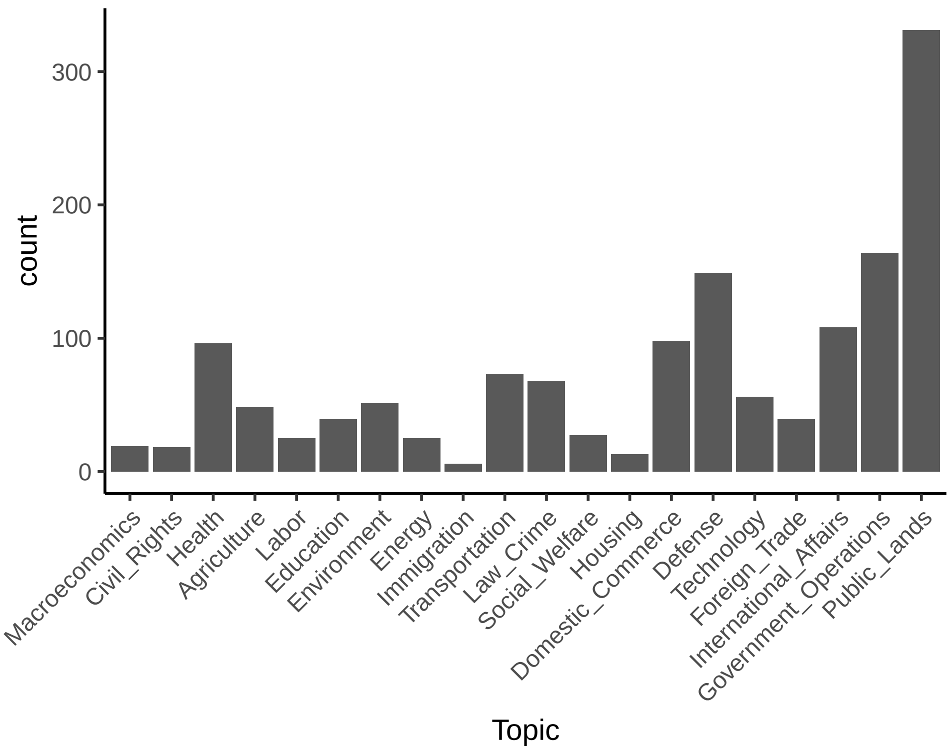
1973-2016 ($n \approx 6000$). First, I scraped the final text of all enacted bills from the https://www.congress.gov/ domain, as well as all metadata available on each bill page. Next, I linked each bill to metadata drawn from metadata drawn from the Congressional Bills Project[9] and from Stewart's Congressional tenure data (Stewart and Woon 2011). Using these metadata, I discarded all so-called "commemorative bills" ($n \approx 1350$), and extracted named entity and relation data from the remaining dataset. I then discarded all named entities that were mentioned only once in their respective documents.[10] Finally, I discarded all bills for which one or fewer unique named entities were recognized (since the density statistic given above is undefined in this case). This process left me with a total dataset of $n \approx 1450$ observations.

Figure 2 gives a visualization from the Enhanced Parnerships with Pakistan Act of

[9]Available at http://congressionalbills.org/.

[10]This heuristic is drawn from the natural language processing literature (see, e.g., Grimmer and Stewart 2013), and is useful in cases where typographical errors are likely to be common.

Figure 3: PAP Major Topic Code Frequency

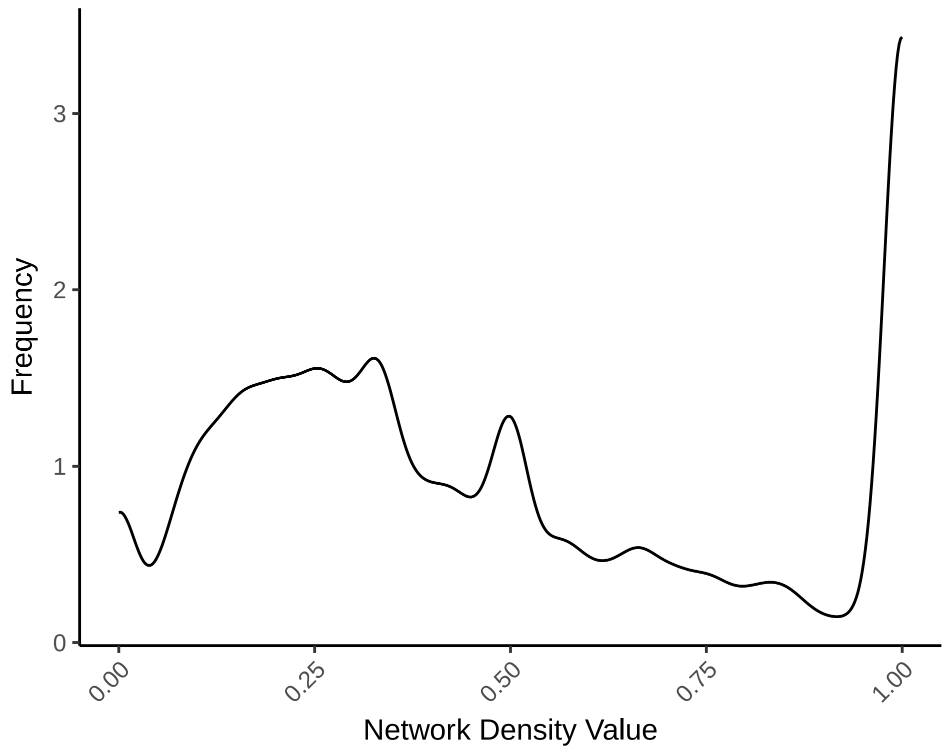


2009[11], a foreign-affairs bill and a typical observation from my dataset. The bill, which provided financial aid to the Pakistani military, primarily links the President, the Government of Pakistan, the Secretaries of State and Defense, and the Appropriations, Armed Services, and Foreign Affairs committees, with other committees and the Pakistani aid fund itself named as peripheral entities. This example highlights both some of the strengths and the challenges in the entity recognition posed as part of this paper. Though the algorithm I describe captures the key actors and the basic set of relationships imagined by the law, it cannot currently associate different terms for the same named entity (e.g. "Pakistan Counterinsurgency Capability Fund"; "Department Defense Pakistan Counterinsurgency Fund"; "Department State Pakistan Counterinsurgency Fund") and sometimes munges distinct terms together (e.g. "Fund Secretary Defense").

Some basic descriptive information on the broader dataset are given in Figures 3 and 4. Across the dataset, all policy agendas codes are represented, with bills falling under the Public Lands topics being the most frequent. As shown in Figure 4, the distribution of the dependent variable is fairly uniform; however, a large number of cases have a density value of precisely one (i.e. all possible ties observed). This feature is problematic for the modeling approach I adopt in the following section, which requires the dependent variable of interest to be strictly between zero and one. As a result, I follow Ferrari and Cribari-Neto (2004) and transform the dependent

[11] S. 1707 (111th Cong.)

Figure 4: Density Value Summary



variable as follows:

$$t(D_i) = \frac{D_i(N-1) - 0.5}{N}$$

This transformation shrinks all points slightly towards the middle of the distribution, with a larger effect on larger points. In informal testing, I also experimented with removing all cases with density values precisely equal to zero or one; however, the results I present below did not vary based on the inclusion or exclusion of these points.

An alternative approach, which I plan to explore further in future work, is to use a zero-or-one-inflated strategy. Since bills with a density value of precisely zero or one are likely to be short bills with relatively little text, they may follow a different data-generating process than longer documents. As a result, we might be interested in modeling the choice between these two document types using a beta/binomial mixture distribution, with separate coefficients for each mixture component (see Ospina and Ferrari (2012) for details).

4.2 Modeling

To gain a sense of the relationships between network density and political variables of interest, I fit a hierarchical Bayesian beta regression model to the network density data described above. As described above, I use the *document* as my unit of analysis; however, I partially pool regression coefficients based on policy area of the law under consideration.[12] Since complexity of the policy area is likely related to legislator willingness to author complex policy stipulations, we might reasonably expect the relationships between political variables and formal allocation of power to vary based on policy area. For example, since many legislators come from a legal background, they might be more willing to expend time and energy authoring bills addressing Law and Crime compared with bills addressing Technology.

As independent variables, I include dummy variables indicating whether the member proposing the bill is part of the same party as the President and whether the member proposing the bill is part of the chamber majority, the logged number of cosponsors of the bill (as a measure of attention to the bill's content), and the DW-NOMINATE score of the proposing member.

The modeling approach I use is largely drawn from standard frequentist formulations of the beta regression (Ferrari and Cribari-Neto 2004). The approach proceeds as follows:

$$\begin{aligned} D_{ij} &\sim Beta(\mu_{ij}, \phi) \\ \mu_{ij} &= f^{-1}(X_{ij}^T \beta_j) \\ \phi &\sim Cauchy(0, 10) \end{aligned}$$

Where X_i is a matrix of predictor variables and D_{ij} is the network density value for the i^{th} document nested in the j^{th} policy area. Here, the Beta distribution is parameterized as in Ferrari and Cribari-Neto (2004)'s treatment, such that if $D \sim Beta(\mu, \phi)$ then $E(D) = \mu$ and $Var(D) = \frac{\mu(1-\mu)}{1+\phi}$, and f^{-1} denotes the inverse logistic link function. Parameterized in this fashion, the model is inherently heteroskedastic; usefully, though, this formulation also allows for direct modeling of the precision term if desired.

As usual in the Bayesian hierarchical modeling framework, I then place shared priors on the regression coefficients:

$$\beta_j \sim MVNorm(\gamma, \Sigma)$$

Where β is a JxK matrix of first-level regression coefficients, with J equal to the number of policy areas and K equal to the number of parameters in the individual-level mean model. Similarly, γ is a $Kx1$ vector of second-level regression coefficients.

[12] As identified by Policy Agendas Project major topic codes assigned by the Congressional Bills Project (http://congressionalbills.org/).

I place priors on the variance-covariance matrix $\Sigma = diag(\nu)\Omega diag(\nu)$ using the strategy suggested by Gelman et al. (2014):

$$\begin{aligned}\nu \sim& Cauchy(0, 2.5)\\ \Omega \sim& LKJ(1)\end{aligned}$$

With ν restricted to the non-negative reals, and with Cauchy parameters selected to produce a vague but mildly informative distribution for regression parameter variances. LKJ refers to the Lewandowski et al. (2009) correlation matrix distribution, which reduces to an identity distribution over covariance matrices with shape parameter equal to 1.

Finally, I place independent priors on the second-level regression coefficients:

$$\gamma_k \sim Cauchy(0, 2.5)$$

With parameters selected here to give vague but mildly informative distributions for regression coefficients on the logistic scale.

4.3 Results

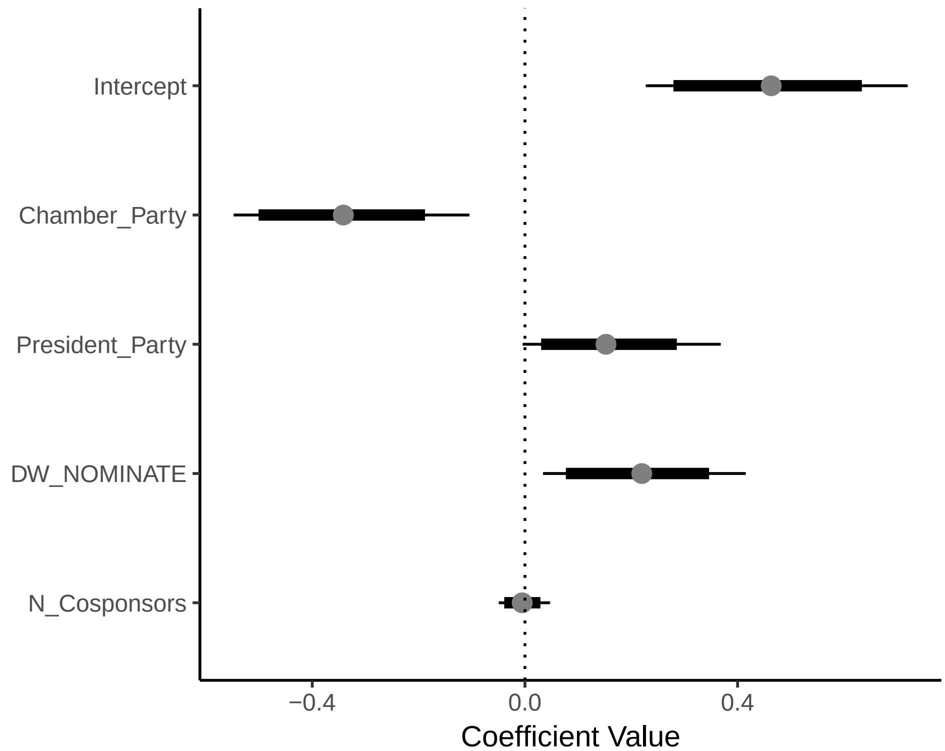
Using the approach described above, I implemented and fit a model to the density data described in Section 4.1.[13] After assessing convergence,[14] I generated a series of visualizations to inspect the results, which are given in Figures. As usual in the GLM setting (Bayesian or otherwise), coefficient values are not directly interpretable on the scale of the dependent variable. In this case, however, the posterior means for the dummy variables on majority party membership (by chamber), shared-Presidential party, and DW-NOMINATE scores correspond to approximately a 30% decrease, a 15% increase, and a 20% increase in observed tie density, respectively, while the coefficient corresponding to the number of cosponsors is estimated roughly at zero.

These signs for the two dummy variables fit with intuitive understandings of the legislative process. Members who belong to the same party as the President are likely more interested in cooperating with the executive branch to create complex bureaucratic structures and "lock in" current preferences (Bendor and Meirowitz 2004), especially if the current majority's downstream electoral prospects are perceived to be weak. The sign on majority in-chamber party, by contrast, likely speaks to a selection effect. Any legislation proposed and by a non-majority member of a chamber that actually passes Congress is likely to be a bill with substantial bipartisan support. As

[13] Using RStan (Carpenter et al. 2016).

[14] In particular, I ran four model chains for 10,000 iterations each, with a 2500 iteration burn-in, and inspected $\hat{R}$ values (potential scale reduction factor; Gelman and Rubin 1992) and effective sample size statistics as given in RStan (Carpenter et al. 2016). For all identified parameters, $\hat{R}$ values were ≤ 1.01 and all effective sample sizes were ≥ 2500, providing no indication that we should reject the null hypothesis of convergence.

Figure 5: Top Level Model Coefficients



Dots give mean values for top-level regression coefficients in the hierarchical beta regression model described in Section 4.2. Thin lines give 95% credible intervals, and thick lines give 80% credible intervals.

a result, any legislation proposed by a non-majority legislature that passes Congress is likely to be a relatively substantial piece of legislation, with more policy complexity than the average bill.

The hierarchical modeling setup described earlier also allows us to explore effect sizes by subgroup (here, PAP topic code). As shown in Figure 6, the difference between bills proposed by members of the majority party in a particular chamber and those proposed by the minority is fairly uniform across all policy areas, reinforcing the selection-related explanation given above. By contrast, the estimated DW-NOMINATE coefficient is much more variable. More conservative members tend to propose bills containing more complex institutional structures in Government Operations, Public Lands, and International Affairs, but most other policy areas show relatively little difference by ideology. These varing coefficients suggest differences in partisan issue priorities, and raise interesting areas for future research.

In contrast to the previous two sets of estimates, the coefficients on attention (number of cosponsors) are much more variable, with substantial posterior mass on both sides of the zero line.[15] Likely, this lack of precision results from measurement error in the dependent variable. As noted earlier, the measurement techniques described in

[15] Coefficient values omitted for readability.

Figure 6: Second-Level Chamber-Majority Coefficient

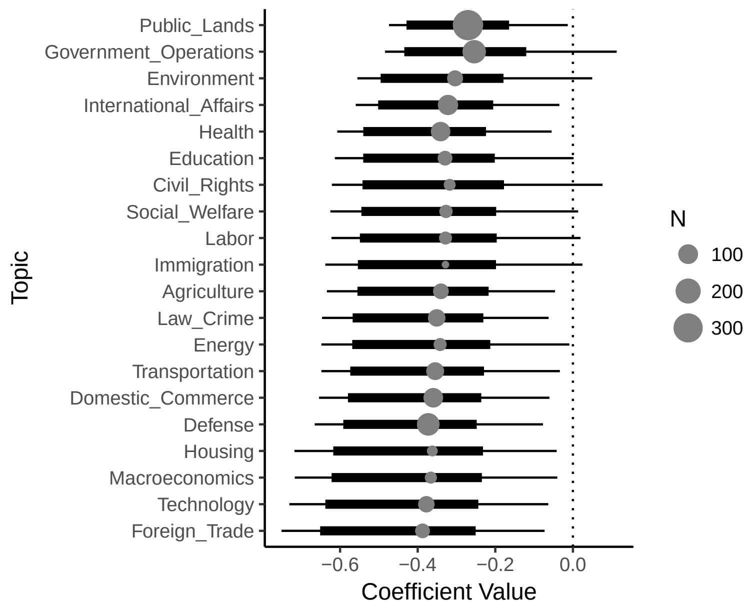
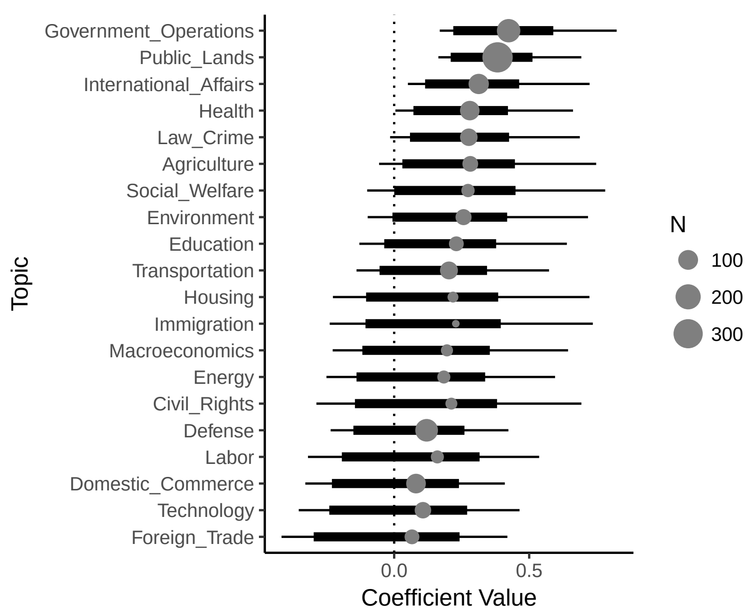


Figure 7: Second-Level DW-NOMINATE Coefficient



Dot sizes are scaled to the number of observations from each group. Thick lines give 95% credible intervals, and thick lines give 80% credible intervals. The baseline category for the Figure 5 is non-membership in the chamber majority party.

this paper are still at a relatively early stage, and the results presented here should be viewed as exploratory. However, this approach offers an encouraging start to a broader analysis.

5 Conclusion

Design of legislation represents a key area of interest for researchers throughout political science. Scholars working in American politics, public policy, and public law alike are all interested in exploring and understanding the processes by which politicians generate legal text, and the consequences (or lack thereof) of their decisions on downstream policy outcomes. As I argue, the existing literature in this area has been constrained by measurement limitations, which have prevented scholars from testing basic hypotheses about the relationships between key political variables and important design features.

In this paper, I conceptualize formal, textual power in *relational* terms, and present a language-based measurement approach designed to capture conceptual scheme. I then apply this approach to an original dataset of American enacted legislation. Generally, I find that legislation proposed by more conservative members of Congress and by members of the chamber minority tend to be more complex (denser) than those proposed by liberal, chamber-majority counterparts. However, these effects vary substantially by policy area, with the bulk of the estimated effect concentrated within a few policy topics. In future work, I plan to extend this measurement scheme to a cross-national sample of legislation, and to other legislative characteristics.

References

Ainsworth, S. H. and Harward, B. M. (2009). Delegation and discretion in anticipation of coalitional drift. *American Politics Research*, 37(6):983–1002.

Baumgartner, F. R., Breunig, C., Green-Pedersen, C., Jones, B. D., Mortensen, P. B., Nuytemans, M., and Walgrave, S. (2009). Punctuated equilibrium in comparative perspective. *American Journal of Political Science*, 53(3):603–620.

Bendor, J. and Meirowitz, A. (2004). Spatial models of delegation. *American Political Science Review*, 98(02):293–310.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., and Riddell, A. (2016). Stan: A probabilistic programming language. *J Stat Softw*.

Dahl, R. A. (1957). The concept of power. *Behavioral Science*, 2(3):201–215.

de Figueiredo Jr, R. J. (2002). Electoral competition, political uncertainty, and policy insulation. *American Political Science Review*, pages 321–333.

Epstein, D. and O'Halloran, S. (1999). *Delegating powers: a transaction cost politics approach to policy making under separate powers*. Cambridge University Press, New York.

Farhang, S. (2010). *The litigation state: Public regulation and private lawsuits in the United States*. Princeton University Press.

Farhang, S. and Yaver, M. (2016). Divided government and the fragmentation of american law. *American Journal of Political Science*, 60(2):401–417.

Ferrari, S. and Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7):799–815.

Franchino, F. (2007). *The powers of the union: delegation in the EU*. Cambridge University Press, New York.

Gailmard, S. and Patty, J. W. (2012). *Learning while governing: expertise and accountability in the executive branch*. University of Chicago Press, Chicago.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian data analysis*. Taylor & Francis, Boca Raton, FL, 3 edition.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472.

Grimmer, J. and Stewart, B. M. (2013). Text as data: the promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3):267–297.

Huber, J. D. and Shipan, C. R. (2002). *Deliberate discretion? the institutional foundations of bureaucratic autonomy*. Cambridge University Press, New York.

Kagan, R. A. (2009). *Adversarial legalism: The American way of law*. Harvard University Press.

Kelemen, R. D. (2011). *Eurolegalism: The transformation of law and regulation in the European Union*. Harvard University Press.

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of multivariate analysis*, 100(9):1989–2001.

McCubbins, M. D. (1985). The legislative design of regulatory structure. *American Journal of Political Science*, pages 721–748.

McCubbins, M. D., Noll, R. G., and Weingast, B. R. (1989). Structure and process, politics and policy: administrative arrangements and the political control of agencies. *Virginia Law Review*, pages 431–482.

McCubbins, M. D. and Schwartz, T. (1984). Congressional oversight overlooked: police patrols versus fire alarms. *American Journal of Political Science*, pages 165–179.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Moe, T. M. (1990a). Political institutions: the neglected side of the story. *Journal of Law, Economics, & Organization*, 6:213–253.

Moe, T. M. (1990b). The politics of structural choice: toward a theory of public bureaucracy. In Williamson, O., editor, *Organization theory: from Chester Barnard to the present and beyond*, pages 116–153. Oxford University Press, New York.

Moe, T. M. and Caldwell, M. (1994). The institutional foundations of democratic government: a comparison of presidential and parliamentary systems. *Journal of Institutional and Theoretical Economics*, 150(1):171–195.

Oosterwaal, A., Payne, D., and Torenvlied, R. (2012). The effect of political disagreement on discretion mechanisms and conditions. *Administration & Society*, 44(7):800–824.

Ospina, R. and Ferrari, S. L. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis*, 56(6):1609–1623.

Roth, D. and tau Yih, W. (2004). A linear programming formulation for global inference in natural language tasks. In *Proceedings of CoNLL-2004*, pages 1–8. Boston, MA, USA.

Stewart, C. I. and Woon, J. (2011). Congressional committee assignments, 103rd to 112th congresses.

Surdeanu, M., McClosky, D., Smith, M. R., Gusev, A., and Manning, C. D. (2011). Customizing an information extraction system to a new domain. In *Proceedings of* *the ACL 2011 Workshop on Relational Models of Semantics*, pages 2–10. Association for Computational Linguistics.