Partisanship or Policy? Understanding Institutional Complexity in American Law

June 22, 2019

Abstract

How do legislators distribute policymaking authority? Formal allocation of power is often framed as a *preference-based* problem. Under unified government, legislators should favor efficient implementing arrangements, while under divided government they should favor complex systems that optimize oversight. By contrast, I argue that this finding is conditioned by *issue importance*. Due to measurement constraints, existing studies focus either on single policy areas or small subsets of "important" legislation, which leads them to ignore this relationship.

To fill this gap, I propose a novel network- and natural language processing-based measurement approach, which I use to measure institutional complexity in all American laws enacted from 1993-2014. As predicted, I find that divided government is only associated with increased complexity for a small subset ($\approx 20\%$) of high-importance laws. Otherwise, this relationship vanishes. This paper therefore contributes to both applied text-as-data scholarship and to the literature on "submerged" bipartisan collaboration in American politics.

1 Introduction

How do legislators allocate policymaking authority? As any would-be lawyer knows, statutes, constitutions, and other formal legal texts establish relationships between actors, describing who can do what, when, and to whom. Debates over these allocation-of-authority decisions have clear policy implications, and often occupy a high-profile position in public discourse. For a concrete example, consider the Patient Protection and Affordable Care Act (ACA). As a core part of Barack Obama's first-term agenda, the ACA represented a sweeping overhaul of the American healthcare system, which expanded availability of public and private insurance plans and empowered regulators to implement an wider set of health insurance regulations. Supporters argued that these changes would protect consumers by empowering expert administrators, while detractors worried that the ACA placed undue authority with a small number of unelected, unsupervised bureaucrats. The discussion surrounding so-called "death panels" offers a immediate example of this latter set of concerns. Though fantastical, the fear that a single government regulator might be given unchecked authority to control end-of-life decision-making for Medicare patients resonated with voters, and represented one of the most durable talking points in the entire ACA debate (Oberlander 2012).

Political scientists and legal scholars have generally explained these allocation-of-authority decisions in *partisan* terms. When the executive's policy preferences are aligned with those of the legislature (and are likely to remain so in the future), legislators tend to pass simple "framework"-style laws involving one or a few implementing actors and straightforward decision-making rules. By contrast, when legislative and executive preferences diverge, legislators tend to favor complex institutional structures with many implementing actors, which provide greater oversight opportunities by fragmenting decision-making authority. This pattern, some authors suggest, helps explain the slow-moving and haphazard policymaking patterns that characterize the American federal bureaucracy. Because divided government is common in the United States, American law often contains fragmented administrative structures that slow bureaucratic action and encourage intervention by courts (see, e.g. Moe

1990b; Moe and Caldwell 1994; Kagan 2009).

However, recent scholarship on minority-party influence over Congressional lawmaking calls this logic into question. Despite increasingly polarized voting patterns in Congress, minority-party lawmakers are often able to influence the content of the final legislative product, both through small-scale amendments and by attaching their proposed bills wholesale as "hitchhikers" to other legislative vehicles (Wilkerson *et al.* 2015; Casas *et al.* 2018). Findings like these motivate the basic puzzle of this paper: if the lawmaking process is more collaborative than previously realized - and has remained collaborative as voting patterns have become increasingly polarized - under what conditions does divided/unified government affect downstream institutional complexity? Are previous findings regarding the relationship between political preferences and complexity of formal institutions generally applicable, or do they depend on the characteristics of the issues and policy areas under consideration?

In this paper, I suggest that the answer to this puzzle lies in the *importance* of the law under consideration. Since institutional design is a costly process, we should only expect lawmakers to exert the effort required to implement their policy preference through legal language on issues which are particularly salient to their public or elite constituents. In other situations, characteristics of the issues and policy problems under consideration should dominate partisan concerns. Due to measurement constraints, previous studies have almost exclusively examined "important", high-salience laws; as a result, the existing literature overstates the influence of partisan divisions on downstream institutional design choices.

To test this proposition, I offer a novel natural language processing-based conceptualization and measurement scheme designed to quickly and easily measure the complexity of institutional structures contained in legal texts. Conceptually, I argue that formal legal texts are best viewed as *relational* documents, which describe which actors can do what, when, and to whom. I then propose and implement a neural network-based approach designed to extract these "implementing networks" from legislative texts, which I validate both qualitatively and in out-of-sample quantitative tests. I then use this measurement approach to study patterns of institutional complexity in a dataset consisting of nearly all enacted American laws passed from 1990-2014, rather than simply the "significant" ones from the period. As predicted, I find that high-importance laws passed under divided government tend to be more complex than those passed under unified government. By contrast, more "everyday" laws passed under unified and divided government are indistinguishable. These findings therefore contribute to both the applied text-as-data literature and to a growing body of scholarship on bipartisan collaboration in the American lawmaking process.

2 The Causes of Institutional Complexity

2.1 Institutional Complexity as Bureaucratic Oversight

When a policymaker writes a formal legal document - such as a constitution, law, or administrative regulation - that policymaker must choose between a range of possible implementing structures. At one extreme, she could create a complicated, interconnected institutional system, which involves many actors in a particular policy area and requires agreement among them to reach a particular policy decision. At the other, she could write a simple "framework" document, which allows one or a few actors to unilaterally make decisions in a given policy area. Farhang and Yaver (2016) succinctly describe this quality - which they term *fragmentation* - as the extent to which a law

> [divides] implementation authority over a larger number of distinct actors, over a larger number of different agencies, and [gives] multiple actors the authority to perform the same function with respect to the same statutory provisions.

The choice between a simple, streamlined implementing structure and a fragmented, diffuse one is consequential. By splitting policymaking authority among several actors, lawmakers reduce implementing efficiency and limit implementer discretion, but offer more opportunities for outside groups to monitor and intervene into the policymaking process. Streamlined implementing structures, by contrast, offer a more efficient policymaking process but fewer oversight opportunities.

To explain variation in institutional complexity, most studies focus on *partisanship-based* explanations. As Moe (1990a; 1990b; 2012) and Moe and Caldwell (1994) argue, complex implementing structures are not *ex ante* desirable, for either policymakers or implementing actors. For policymakers, complicated power-sharing arrangements slow implementation and promote gridlock, while for implementers these structures warp core agency "missions" and create new priorities and failure points through which they might be held accountable (Wilson 1989). Worse, complex implementing structures are costly for lawmakers to design. If a lawmaker wants to create a complicated institutional structure involving many implementing actors, she will need to devote a substantial quantity of time and attention to the institutional design process. For example, she might need to carefully research the existing legal and administrative landscape, interview expert interest group and agency leaders, and design incentives, evidentiary rules, and decision-making procedures to encourage bureaucrats to reach favorable policy decisions. Thus, at least in the abstract, both policymaker and implementer should prefer "framework" legislation that offers one or a few implementing actors substantial discretionary authority.

Here, Moe's "politics of structural choice" intervenes. Whenever policymaker and implementer possess different substantive preferences - for example, during periods of divided government - institutional complexity becomes more attractive for legislative actors.[1] By creating overlapping decision-making processes that involve a larger number of implementers, lawmakers can slow the rollout of undesirable policy programs and generate new opportunities for "fire alarm"-style oversight from interest groups and the courts (McCubbins and Schwartz 1984; Kagan 2009; Farhang 2010). Institutional complexity, from this standpoint, provides a means for legislators to passively check the administrative state, even after members of the current majority coalition lose control of formal oversight tools.

This framework produces a natural set of predictions. Based on this logic, when the

[1] Ting (2003) makes similar claims with respect to redundant bureaucratic structures, which can help lawmakers to achieve more favorable policy outcomes when their priorities are not aligned with those of bureaucracy.

executive and legislative branches share similar interests, the implementing structures contained in a given legal document should be *simpler*. In other words, under unified government, enacted laws should name fewer implementing actors, and the decision points those laws describe should require input from fewer decision-makers. Of course, policy and political context matter too; for example, majority-party lawmakers might favor simpler implementing structures under unified government, but might nevertheless create complicated decision-making structures in defense or macroeconomics laws, which frequently address policy problems that cut across many administrative departments. But, based on this framework we should expect executive/legislative preference disagreements to produce more complex decision-making structures at a rate that is roughly constant no matter the policy area or political context.

These predictions have received at least suggestive support from a variety of empirical studies, which cut across national context, policy area, and conceptualization and measurement approach. Most prominently, Farhang and Yaver (2016)'s study of historically "important" legislation reports the positive relationship between divided government and fragmentation predicted by this framework. A large body of scholarship on specificity of legal and policy language also reports analogous findings, in the US Congress (Epstein and O'Halloran 1999), the European Community (Franchino 2004, 2007), and - at least for sufficiently professionalized legislatures - in American state-level lawmaking (Huber *et al.* 2001; Huber and Shipan 2002).[2] As Farhang and Yaver (2016) argue, fragmentation is distinguishable both theoretically and empirically from specificity, in the sense that a statute containing highly specific policy language can nevertheless locate all decision-making authority in a single actor. However, since institutional structures that encourage fragmentation - such as reporting consultation, and appeal requirements - form a part of the coding schemes used by Epstein and O'Halloran (1999) and Franchino (2007), these studies offer some indirect

[2]In particular, Huber and Shipan (2002) find a positive relationship between divided government and statute length (a proxy for regulatory detail) in states with sufficiently professionalized legislatures, and a null relationship otherwise.

additional support for the partisanship-based framework I describe in this section.

2.2 Legislative Importance and the Costs of Complexity

Though appealing and influential, this partisanship-based empirical framework is also limited. Despite growing partisan polarization in Congressional voting patterns, recent studies have found that minority-party lawmakers retain substantial influence over the *content* of legislation. For example, Wilkerson *et al.* (2015) demonstrate that, though the Affordable Care Act received zero Republican votes, as much as 25% of its substantive content was drawn from bills originally proposed by Republicans in the same Congressional session. Similarly, Casas *et al.* (2018) find that majority-party "hitchhiker" laws - or, laws which were proposed as independent bills but enacted as part of an unrelated laws - are only 25-50% more likely than their minority-party counterparts to pass, compared with a 200-300% estimated difference for majority-party complete bill proposals. Though findings like these tell us relatively little about the characteristics that make minority-party proposals successful, they do suggest that minority-party lawmakers exert substantial influence over the content of legislation, even under unified government.

Findings like these suggest an important additional factor for the downstream design of legislation: namely, the *importance* lawmakers assign to the issues contained in a given law. For the purposes of this paper, by "important" I refer to those issues that capture publicly-expressed attention from lawmakers' public and elite constituents and partisan stakeholders. Major "landmark" laws like the 2017 Tax Cuts and Jobs Act (Pub. L. 115-97) clearly meet this standard, while narrower laws and "hitchhiker" bills - which pass as lower-profile attachments to larger laws - usually do not. American politics scholars have long noted the interaction between issue importance, partisan balance of power, and bill success; for example, enactment rates for "important" bills decline under unified government, while their lower-profile counterparts pass at roughly equal rates (see, e.g. Coleman 1999; Howell *et al.* 2000). All else equal, we should expect an analogous pattern to hold with respect to the content of legislation. Lawmakers should be more willing to create *more complex* institutional structures in more important laws, and the relationship between divided government and institutional complexity should be *magnified* when lawmakers believe the issues contained in a law are more important.

The logic behind these expectations is straightforward. As I describe in the previous section, institutional design is a costly process, and complex decision-making arrangements are more labor-intensive for lawmakers to design than simple ones. As a result, *institutional complexity is not equally appealing in all situations.* Rather, lawmakers will be more willing to exert the energy required to create a complex implementing structure to the extent that they view the policy problem addressed by a given law as an *important* issue. Of course, issues can capture public attention for a variety of reasons. For example, periodic negotiations like those over various Farm Bills create predictable spikes in public attention to agricultural policy, while exogenous shocks like the 2008 Financial Crisis create more erratic spikes in public attention to banking and macroeconomics. But, whatever the source, increased public attention to a problem should prompt politicians to examine that problem more closely, and should render lawmakers more willing to craft complex implementing solutions.

Crucially, we should also expect the "importance" of the policy problems contained in a law to condition the relationship between partisanship and institutional complexity. This idea - which mirrors a similar conditional relationship between partisan conditions, importance, and specificity of American state-level legal language reported by VanSickle-Ward (2014)[3] - again relates to the costs of institutional design. When addressing lower-priority issues, politicians have fewer incentives to carefully craft implementing structures that implement their preferred policy outcomes. In these cases, legislators should be more willing to accept content proposals (and accompanying institutional design suggestions) from minority-

[3]VanSickle-Ward (2014) finds that, when considering "important" issues, lawmakers are more likely to craft more "specific" laws under divided government than unified government. But, when considering less consequential issues, this relationship vanishes. Since VanSickle-Ward's dependent variable is *specificity* rather than *fragmentation*, her study is distinct from the one contained in this paper. However, I expect a broadly similar mechanism to operate in this setting.

party lawmakers, which mutes the effect of executive-legislative preference disagreements on downstream institutional design choices. By contrast, when addressing high-importance policy problems the relationship between preference disagreements and downstream institutional complexity should be large and positive, since politicians should be more willing to use institutional complexity as a tool of bureaucratic control.

2.3 Measurement Limitations as Theoretical Constraints

Before continuing, it is worth pausing to consider why "importance"-type effects have been largely ignored in the existing literature. After all, an extensive body of work in the bounded rationality (Baumgartner *et al.* 2009; Baumgartner and Jones 2015) and broader public policy literatures (e.g. May and Jochim 2013) suggests that we should only expect legislators to actively reconsider existing institutional and policy regimes response to major exogenous shocks. We should therefore further expect partisan preferences to have a disproportionate impact on the content of legislative proposals when lawmakers are considering high-importance problems, as opposed to their more day-to-day counterparts. However, with the notable exception of VanSickle-Ward (2014) - who focuses on state-level legislation within a single policy area - few existing studies focus on this relationship.

The reason for this gap, I suggest, is simple: *measuring text-based distribution of authority is difficult.* Even for experts, parsing legal texts is difficult and labor-intensive. To underscore the extent of this measurement challenge, consider Farhang and Yaver (2016)'s study of fragmentation patterns in federal US legislation. In their paper, the authors read and coded some 24,000 pages of legislative text in order to produce data on some 366 "historically significant" laws passed from 1947 to 2008. Though heroic, hand-coding protocols like these are impractical to apply beyond the relatively small corpus these authors examine. Huber and Shipan (2002)'s word-count method - which treats statute length as a proxy for the precision and complexity of that statute's legal and institutional structures - does offer a more scalable alternative (see also Clinton *et al.* 2012). However, because different policy areas and legal contexts possess different drafting standards and preexisting associated bodies of law, this method is inapplicable for comparisons across different policy domains.[4] As a result, most existing studies of formal institutional design patterns have been forced to limit the scope of their work, either by focusing on single policy areas (e.g. Huber and Shipan 2002; VanSickle-Ward 2014) or (more commonly) by focusing on high-salience, "important" legislation (e.g. Epstein and O'Halloran 1999; Farhang and Yaver 2016).[5]

Limitations like these are problematic in at least two respects. First, from a descriptive standpoint, the labor-intensive nature of textual data collection efforts means that we know very little about the content of most laws, even in an institution as well-studied as the US Congress. Most laws enacted by Congress are not the kinds of transformative legislative actions that receive the bulk of scholarly attention. For example, take Mayhew (1991)'s list of "significant" legislation, an extended version of which forms the sample used by Farhang and Yaver (2016). For the period from 1993-2014 (103rd-113th Congresses), Mayhew's list contains some 128 pieces of legislation, compared with 4,609 total public laws passed during the period.[6] Of course, not all of these enactments were impactful; approximately one-quarter (1,126) were relatively trivial "commemorative" laws, which established monuments or symbols designed to commemorate noteworthy people or places in US history.[7] The remaining three-quarters are highly heterogeneous, ranging from technical, procedural laws[8] to important laws that transformed a particular policy domain. For a concrete example, consider the Enhanced Partnership with Pakistan Act of 2009 (Pub. L. 111-73), which authorized some $1.5 billion in yearly defense-related aid to the government of Pakistan from 2010-2014.

[4]Even within a single policy domain, this method is likely imprecise. As Huber and Shipan acknowledge, statutes can be longer because they describe a single policy program with greater specificity, because they describe a larger number of programs, or because they contain more "filler" language with little legal impact.

[5]Anastasopoulos and Bertelli (2017) offer a promising supervised learning-based alternative designed to address this measurement problem. However, their method relies on the availability of hand-coded training data. Because training data of this kind is labor-intensive to collect, obtaining the necessary training data is likely unrealistic in many research settings.

[6]See Mayhew's list for details.

[7]As defined by the Congressional Bills Project.

[8]E.g. Pub. L. 108-306, "To provide an additional temporary extension of programs under the Small Business Act and the Small Business Investment Act of 1958 through September 30, 2004, and for other purposes".

Though not significant enough to make a list like Mayhew's, the Enhanced Partnership with Pakistan Act nevertheless involved substantial financial outlays and represented an important shift in US-Pakistan relations. Unfortunately, measurement limitations like those I describe above mean that we know almost nothing about the prevalence or content of these important but not "historically significant" laws.

Second - and perhaps more importantly - measurement and data limitations like those I describe inhibit our ability to test key outstanding theoretical claims. The focus on high-importance legislation common in existing work would be less problematic if importance was unrelated to downstream institutional design choices, or to key predictor variables such as the presence of divided or unified government. Unfortunately, as I describe in the previous section, features of the policy areas and policy issues addressed by a given law are likely both to directly influence its institutional content and to condition the relationships between standard preference-related variables and the institutional structures favored by legislative actors. As a result, without a more scalable measurement approach, we cannot be sure whether patterns described in existing work generalize beyond historically "important" legislation examined in existing work.

3 Using Implementing Networks to Measure Institutional Complexity

To address this measurement challenge, in this section I present a novel network- and natural language processing-based conceptualization of formal institutional power, which I use to develop a scalable measure of formal institutional complexity. In social science research, a common way to study the power relationships contained in a document is to study the *relational* statements that document contains. For example, Franzosi *et al.* (2012) use a semi-automated method to study portrayals of agency and victimhood in newspaper accounts of lynching episodes in the American South. Among other findings, their analysis reveals a surprisingly strained relationship between law enforcement officials and white lynching mobs, who targeted sheriffs and deputy sheriffs for violence in nearly half of all lynching episodes. In political science, a notable example of this kind of approach is GDELT (Leetaru and Schrodt 2013), which mines news accounts for subject/action/object triples corresponding to international events.

This approach is similarly helpful in the legal setting. Among other functions, legal texts articulate power relationships, describing who can do what, when, and to whom. As a result, a reasonable way to conceptualize a law, contract, or constitution is to view it as a *network* of institutional relationships. The "nodes" in this network represent actors involved in the execution of powers outlined in the text, while the "edges" represent the relationships between them. Depending on the scope of the document, these relationships might be directed or undirected, and might include simple connection types or more complex ones. To take an example familiar to most Americans, the US Constitution provides the House and the Senate joint authority over "legislative Powers herein granted"[9] (an undirected relationship) while empowering the Senate alone to offer "Advice and Consent"[10] to the President over treaties, and judicial/executive appointments (a directed relationship). By collecting and combining all such entities and relational ties, we can construct an "implementing network" for the US Constitution, which describes the set of formal institutional relationships envisioned by the document. We can then use this representation to construct measures of downstream quantities of interest such as the overall complexity of the network, which I turn to in §4.

Besides its conceptual and measurement advantages, a relational conceptualization of formal power also helps to pinpoint the technical challenges involved in studies of this kind. At their most fundamental level, networks are constructed from *entities* (the actors involved in the network) and *edges* (the ties between them). Treating formal legislative texts as implementing networks implies two tasks: in particular, we need to identify the *set of implementing actors* named in a particular legal document, and the *relationships* between those

[9]U.S. Const. Art I, §1.
[10]U.S. Const. Art II, §2.

actors. As I show in the following subsections, both of these quantities can be extracted in a straightforward, scalable fashion using natural language processing methods.

3.1 Entity Extraction

To construct an implementing network for a particular legal text, the first challenge is to identify the set of actors named by that text. In political science, perhaps the most common approach to this problem is a dictionary-based system (see, e.g. Leetaru and Schrodt 2013), in which the entities of interest are pre-identified using a dictionary generated by expert researchers. This approach produces few false negatives, but misses a large number of items of interest, since generating a comprehensive named entity dictionary is impractical in most situations. By contrast, general-purpose machine learning approaches - which rely on lexical and grammatical information to make tagging decisions - usually capture most entities of interest, but will also include irrelevant entries such as the names of people or places (Manning *et al.* 2014). For the purposes of this paper, I therefore opt for a customized machine learning approach, which splits the difference between these two alternatives by leveraging expert-generated lists of government entities to train a specialized named entity recognition model.

To collect training data for this method, I used a three-step procedure. First, using Wikipedia and US government sources, I built a custom dictionary ($n = 1,346$) of common federal US government entities, which includes executive-branch agencies and departments, Congressional committees, and government-sponsored enterprises.[11] Second, I

[11] In particular, I first scraped names from usa.gov, the Federal Register, or one of five Wikipedia sources: specifically, the lists of federal agencies, defunct federal agencies, House committees/subcommittees, Senate committees/subcommittees, and joint committees. I then removed common prefixes and suffixes from these items (e.g. "United States"; "USA"), and stripped names of states and national governments (e.g. "Texas"; "California"; "Federated States of Micronesia") from the list. As an additional quality control measure, an undergraduate research assistant read a random sample of 1,000 sentences, and supplemented this list with a series of additional missing items.

scraped, cleaned,[12] and split into sentences[13] all non-appropriations bills included in legislation enacted from 1993-2014.[14] Third, for each sentence, I conducted a simple string search for each entity contained in the entity dictionary, and tagged any entities found using a standardized (I)nside/(O)utside/(B)eginning (IOB) tagging scheme (see Table 1 for an example and Ramshaw and Marcus (1999) for discussion).[15] This process left me with some 90,711 sentences containing at least one named entity, which formed the training set for the tagging model I employed.

As a tagging model, I use a long short-term memory (LSTM) neural network, which is a common approach in modern natural language processing work.[16] Broadly, LSTMs are a type of recurrent neural network, which allow the tagging decision for each individual token to be influenced both by characteristics of the token and by characteristics and tagging predictions for "adjacent" data points. This approach creates a recursive, context-sensitive prediction structure ideal for analyzing sequentially-organized data. For the purposes of this paper, I specifically rely on the bidirectional CRF-LSTM neural network architecture proposed by Lample *et al.* (2016) and Ma and Hovy (2016).[17]

How well does this custom-trained LSTM approach perform? Since the purpose of using a machine learning approach for named entity recognition is to capture named entities not already known to the researcher, the most relevant (and stringent) performance test

[12] In order to maintain parity between the LSTM training set and the testing and prediction sets, I used identical preprocessing steps for both training and prediction. See §4.1 for a description of the cleaning procedure used for each document.

[13] Using the pretrained Punkt sentence tokenizer, available via NLTK.

[14] As defined using Casas *et al.* (2018)'s data. This dataset includes both enacted legislation and "hitchhiker" bills (independently-proposed bills included as amendments to other enacted legislation).

[15] Since some named entities are substrings of others - for example, compare "Secretary of Defense" with "Assistant Secretary of Defense" - a naive implementation of this scheme might create incorrect tagging choices. For example, if the string "Assistant Secretary of Defense" was tagged before "Secretary of Defense", both "Assistant" and "Secretary" would receive "B" tags. To avoid this problem, I configured the tagging protocol to only tag tokens that had not already received a named entity tag. I then ordered the entity dictionary from longest to shortest, to ensure that longer named entities would be tagged first.

[16] E.g. language modeling (Sundermeyer *et al.* 2012) and part of speech tagging (Huang *et al.* 2015; Plank *et al.* 2016).

[17] As implemented in Tensorflow and Python in the tf_ner library. This implementation is slightly different from the one outlined in the two papers I cite in-text; for details, see the accompanying documentation. See Appendix A.3 for details regarding parameter settings and training.

Table 1: Sample training example

Token	Tag
Funds	O
herein	O
appropriated	O
to	O
the	O
Department	B-MISC
of	I-MISC
Defense	I-MISC
for	O
construction	O
shall	O
be	O
available	O
for	O
hire	O
of	O
passenger	O
motor	O
vehicles	O
.	O

Sample training example, formatted according to a modified version of the CoNLL2003 format. Military Construction Act 1992 §102. For original text see the corresponding congress.gov page.

would be one in which we assess the model's ability to recover *unseen entities* not available during training. In order to assess the model's performance in this scenario, I therefore conducted a five-fold cross-validated predictive accuracy study. First, I randomly split my entity dictionary into five equally-sized sets of entity names. For each set, I then identified all sentences exclusively containing entities from the set in question, and used these sentences to form a held-out test set. I then trained a model using sentences containing entities from the remaining four groups, and calculated precision, recall, and F1 scores for the held-out test set.[18] Finally, I repeated this process for each group, and averaged the performance statistics to produce the final set of results.

Assessed according to these metrics, the LSTM model I employ achieved a cross-validated F1 score of 0.817, with a cross-validated precision of 0.904 and recall of 0.743. These values are comparable to or higher than those reported elsewhere in the literature; for example, Augenstein *et al.* (2017) report an overall F1 score of 0.507 for entity recognition on unseen named entities, averaged across three modeling approaches and 19 standard benchmarking datasets. For the standard CoNLL datasets - which contain newswire text that is relatively similar to the language I examine in this paper - these authors report F1 scores of 0.844 and 0.871. Overall, then, the performance results I report match state-of-the-art performance on standard testing datasets, and represent a strong basis from which to work.

3.2 Relation Extraction

The second step to constructing an implementing network is to identify the *relationships* between actors named by a text. For the purposes of this paper, I focus on *co-occurrence* relationships, which I define as instances in which two actors are assigned to implement the same policy program. This definition is similar to the one used by Farhang and Yaver (2016), and offers both theoretical and practical advantages. As I argue in previous sections,

[18] Defined as $F1 = \frac{2PR}{P+R}$, $P = \frac{TP}{TP+FP}$ and $R = \frac{TP}{TP+FN}$, where P and R denote precision and recall, $T(P|N)$ denote the count of true positive/negative examples correctly classified and $F(P|N)$ denote the counts of true positive/negative incorrectly classified.

institutional structures are complex to the extent that they include more actors and create a denser network of decision-making structures connecting those actors. Involving more actors in a policy decision creates more opportunities for "fire alarm"-style oversight and more decision points that outside interest groups can use to influence the policymaking process. The key quantity of interest is therefore the *number* of actors involved in a particular policy decision and the *number* of ties that connect them. Details about the nature of these ties - such as the hierarchy of the two entities, or the role assigned to each - are informative, but less significant. So long as we can be reasonably confident that each policy decision under consideration is roughly comparable in its substantive scope, simply counting the number of actors involved in a law's policy decisions offers useful information about the complexity of the institutional relationships that law contains.

Fortunately, drafting guidelines for American legislation make these kinds of co-occurrence relationships relatively easy to identify. As noted in the drafting guide for the US Consolidated Code, the "basic unit" of federal US law is the *section*.[19] According to the Code, *sections* "shall contain, as nearly as may be, a single proposition of enactment".[20] 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.[21] Since sections in modern legislative texts are demarcated using consistent formatting rules, drafting rules like these offer a useful way to identify instances in which decision-making authority is split between one or more actors.

Though effective for this paper, this approach is clearly not applicable for all potential research questions. Because the co-occurence networks I construct tell us relatively little about the nature of the relationships between actors, this method cannot be used to determine which entities have oversight or veto power over others, or which decisions require formal approval (as opposed to simple advice consultation) from multiple entities. Rather,

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

[20] 1 U.S.C. §104

[21] As noted in §5, during preprocessing I remove some sections that obviously violate this assumption, such as short titles, findings of Congress, and "Table of Contents" sections.

the co-occurence networks this method produces are most useful for summarizing the broad character of the institutional structures contained in a law, through statistics which describe that structure's complexity, size, or density.

3.3 Constructing Implementing Networks

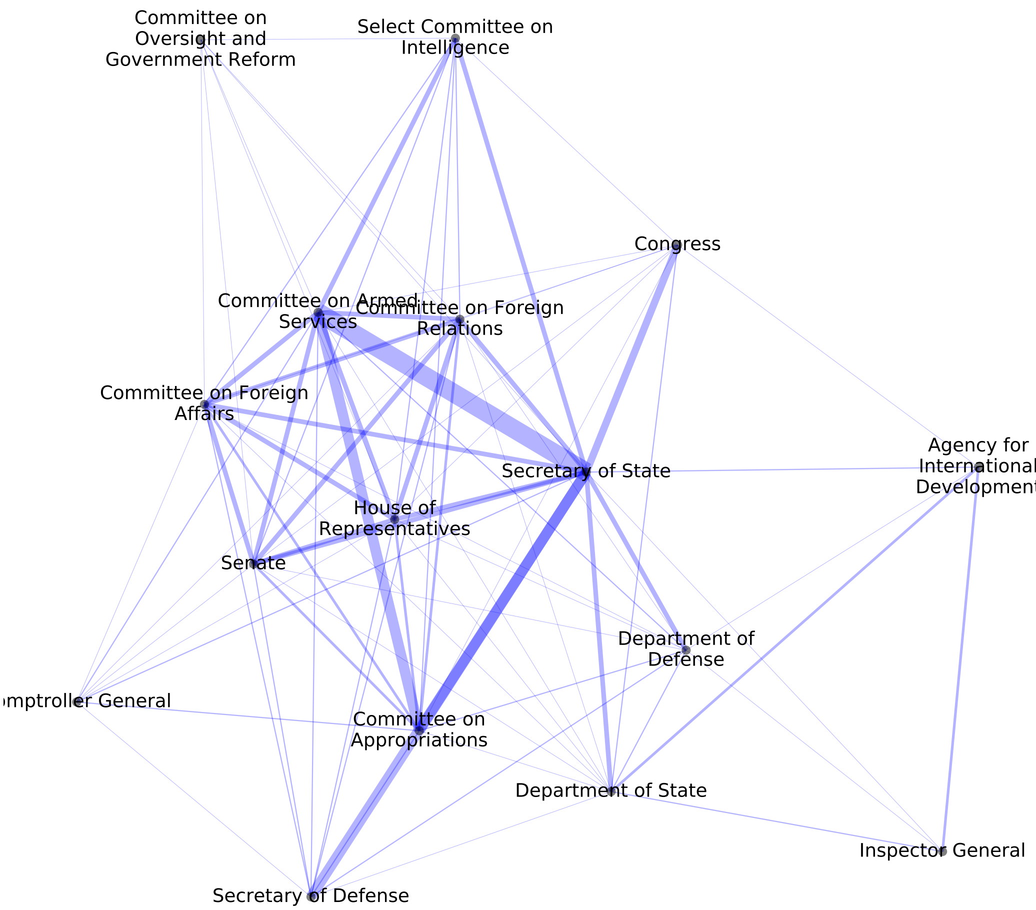
Put together, the named entity and relation extraction approaches I describe offer a simple three-step implementing network construction procedure. First, I split each text into sections[22] and extracted entities from each section using the LSTM model I describe in §3.1. Second, I drew an edge between any set of entities that co-occur in a given section. This process yields a distinct implementing network for each law, which can be used both to describe the law's institutional content and to calculate document-level quantities of interest like institutional complexity.

To illustrate this process, I provide a sample output from the Enhanced Partnership with Pakistan Act of 2009 (Pub. L. 111-73). The Enhanced Partnership with Pakistan Act was a relatively straightforward foreign aid law intended to provide military and developmental assistance to the Government of Pakistan. The law authorized the President to provide $1.5 billion in non-military aid from 2010-2014, and provided additional military aid conditional on a certification process implemented by the Secretary of State.[23] Unusually for a defense-oriented law, the Enhanced Partnership with Pakistan Act gave the State Department substantial authority over defense-related aid allocations (Epstein and Kronstadt 2013). As shown in Figure 1, these features are clearly visible in the law's implementing network. The law contains a central cluster consisting of the Secretary of State, the Secretary of Defense, and several Congressional actors. Quantitative assessments of node importance reinforce this visual message; as measured by eigenvector centrality, the Secretary of State is the most central actor in this network (eigenvector centrality of 0.45), followed by the

[22]Using the constitute_tools regular expression parser. This parser separates each document according to a given set of organizational headers (e.g. titles; sections), while maintaining the internal hierarchy of each document. See Appendix A.1 and A.2 for details and sample parsed text.

[23]Enhanced Partnership with Pakistan Act of 2009, §203.

Figure 1: Implementing network, Enhanced Partnership with Pakistan Act of 2009.



Line density is approximately proportional to the number of ties between each node. Node placement is random, but is loosely related to node centrality.

the Committee on Armed Services (0.39), the Committee on Appropriations (0.32), and the Committee on Foreign Relations (0.29). These figures roughly track with qualitative summaries of the law's content, lending this representation a substantial degree of face validity.

4 Institutional Complexity in American Law

Throughout this paper, my key theoretical contention is that the relationship between executive/legislative preference disagreement and institutional complexity should be positive on publicly visible, "important" legislative proposals, and null otherwise. To investigate this hypothesis, I used Casas *et al.* (2018)'s data to construct a comprehensive dataset of bill texts (and accompanying metadata) included in laws passed from 1993-2014 ($n = 6,109$).[24] This dataset includes both independently-enacted bills and separately-proposed "hitchhiker" laws that were eventually included in enacted legislation. As Casas *et al.* (2018) demonstrate, many bills proposed by members of Congress that fail to pass are later included as "hitchhikers" onto future enacted bills.[25] Because these "hitchhiker" bills are drafted and proposed separately from the laws in which they are eventually included, we should expect their content to be primarily influenced by the characteristics of lawmakers who originally proposed them. In order to enable these kinds of comparisons, I therefore isolated the text of each "hitchhiker" bill from the enacted law in which it was eventually included, and treated these "hitchhikers" as distinct units.[26]

[24]This dataset excludes two types of laws: specifically, "commemorative" laws, and appropriations and budget laws. As defined by the Congressional Bills Project, "unimportant" laws include "commemorative" laws and laws which transfer small quantities of land between government entities. Since these laws are not intended to affect the structure of the administrative state, they are unlikely to include allocation-of-authority language. Appropriations and budget laws are also unlikely to include allocation-of-authority language, but for different reasons. Because these laws are primarily designed to disburse funds to various agencies and governmental entities, most of their language is devoted to detailed descriptions of funding decisions, rather than to the powers and duties of governmental entities. As a result, I retain "hitchhiker" bills that eventually became part of an appropriations or budget law (since these "hitchhikers" are usually unrelated to the budgeting and appropriations processes), but remove appropriations and budget laws themselves from the dataset.

[25]Usually, as a separate title or chapter of the enacted law.

[26]In particular, for bills identified by Casas *et al.* (2018) as "hitchhikers", I compared each section of the hitchhiker to each section of the final "target" law. I then removed all sections from the target law that shared at least 95% of their tokens with a section from the hitchhiker and differed in length by no more than

Before proceeding, I conducted some simple preprocessing steps. For each text, I first stripped headers, footers, and editor's notes (e.g. date of passage; legislative history; transcription notes). I then split each law text into sections, and removed sections containing tables of contents, short titles, and related non-legally binding language.[27] Finally, for all remaining sections I removed section titles and law common names from each section's text.[28] Because law names often contain the names of the administrative agencies or institutions whose jurisdiction they affect, these names are a common source of false positive named entity results.[29]

4.1 The Dependent Variable: Network Complexity

As I describe in §2.1, the formal implementing structures contained in a given text are *complex* to the extent they grant a larger number of actors overlapping responsibility over execution of a particular policy program. In the network context, the most plausible operationalization of this quantity is the *average degree* of the network. This quantity can be interpreted as the average number of instances in which a given actor is coinvolved with other actors in the implementation of a particular policy program.[30] To calculate this value, I use the procedure I describe in §3 to extract a distinct implementing network for each law, and calculate an average degree statistic for each law.

Some basic descriptive information for this variable is provided in Figure 2. As this plot suggests, the average degree measure is both widely dispersed and zero-inflated, with a maximum of 639 and some 1,174 observations (approximately 20% of the overall dataset) with an average degree value of zero. Moreover, some 1,813 laws (approximately 30% of the dataset) contain zero named entities, leaving them with an undefined average degree value

5%.

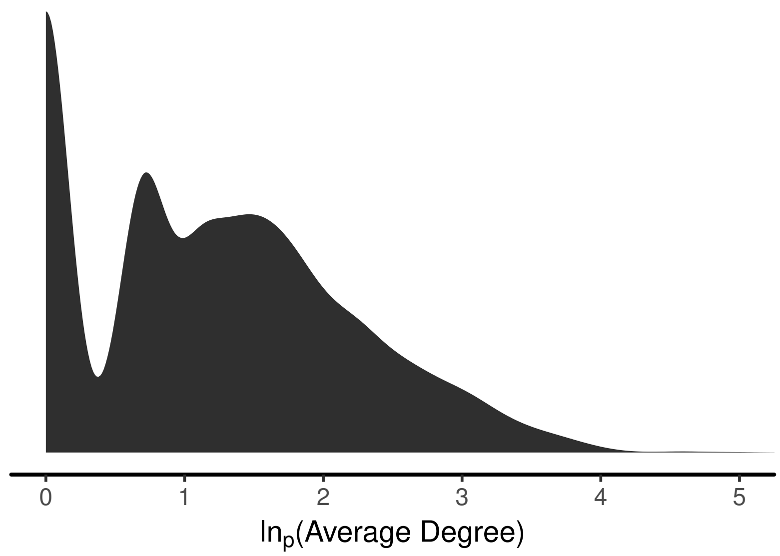
[27] In particular, I removed all sections containing any of the terms "Short Title", "Table of Contents", "Finding", "Purpose", "Definition".

[28] Using the Cornell Legal Information Institute's list of law common names.

[29] For example, the Poison Control Center Enhancement and Awareness Act (Pub. L. 106-174).

[30] Formally, the degree of a node in a real-valued network refers to the sum of all edge weights connected to that node. The average degree simply represents the degree value averaged over all nodes in the network.

Figure 2: Density plot, per-law average degree value.



Density plot, log-plus-one transformed average degree value ($n = 4,296$). Bills with undefined average degree values not displayed.

and reducing the final dataset to $n = 4,296$ observations.

Both of these features are unsurprising from a theoretical standpoint. As I describe in §2.3, not all laws alter the powers or restrictions of government actors. Rather, a non-trivial proportion of the dataset consists of short, technical bills that do not alter the jurisdiction of any actor.[31] Bills of this latter type - which usually contain zero or a small number of named entities - likely follow a different data-generating process than the one I focus on in this paper, which I address from a technical standpoint in the modeling discussion in §4.4.

4.2 Predictor Variables

Based on the expectations I describe in §2.2, my key predictor variables are the presence of executive-legislative preference disagreement and the public importance of a given law. To operationalize executive-legislative preference disagreement, I use a simple binary indicator,

[31] E.g. Pub. L. 108-306, "To provide an additional temporary extension of programs under the Small Business Act and the Small Business Investment Act of 1958 through September 30, 2004, and for other purposes." This bill simply extends authorization for existing provisions of the Small Business Investment Act of 1958, and therefore provides no modifications to existing administrative jurisdiction.

which consists of a dummy variable indicating the presence or absence of divided government at the time that the law in question was passed.[32] For law importance, I follow Volden and Wiseman (2014) and use a binary indicator denoting whether the law in question was mentioned in the CQ Almanac's year-end summary of Congressional activity. Though coarse, this measure offers a useful indicator for the "importance" of the policy concerns contained in a given law, at least to "insiders" and Congress-watchers who are engaged with Congress's activities. Finally, to capture the conditional relationship I posit in §2.2 I include an interaction between these two indicators. Broadly, I expect the divided government coefficient to be approximately zero and the interaction coefficient to be positive. By contrast, the coefficient on the "importance" variable should be positive in all situations, since legislators should be more willing to invest in institutional complexity on issues that receive heightened public attention no matter the political circumstances.

Besides these variables, I also included a series of other law- and individual-level controls. For the law-level variables, I first included a random intercept term corresponding to the Congressional Bills Project's 20 law-level policy codes. This variable is intended to capture policy area-specific institutional design patterns, and should be larger for more complex policy areas (e.g. macroeconomics or defense) and smaller in simpler ones (e.g. public lands). I also included a binary variable indicating whether the bill in question was a "hitchhiker" or an independently enacted law, and a predictor corresponding to the square root of the number of cosponsors in each law. Since hitchhikers are by definition subsets of other laws, we should likely expect these bills to be shorter and simpler than independently enacted laws on average, leading to a negative coefficient estimate.[33] Cosponsorship, by contrast, is intended to capture the quality and salience of a given law within Congress. I broadly expect the relationship between cosponsorship levels and institutional complexity to be positive. However, since cosponsorship incentives are heterogenous across law types and

[32]See Appendix B for descriptive statistics on all predictor variables.

[33]Excluding "hitchhikers" entirely from the model also leaves other model estimates substantively unaffected. See Appendix C.2.4 for details.

policy areas, the magnitude of this coefficient may not be large.[34]

For individual-level predictors, I included an binary variable indicating whether the proposing member was part of the chamber majority, and continuous variables corresponding to the DW-NOMINATE and squared DW-NOMINATE score of the proposing member. Since members of the chamber majority by definition control at least one chamber of Congress, they have a greater ability to enact ambitious, transformative laws, which are likely to require more complex implementing structures. For the DW-NOMINATE variables, expectations are less clear. All else equal, both liberals and conservatives have incentives to use institutional complexity as a policymaking tool. Liberals, for their part, are likely more willing to propose bills with more ambitious policy goals, which tend to require more complex implementing structures. Conservatives, by contrast, tend to be skeptical of the administrative state, and thus might be willing to use institutional complexity as an oversight mechanism. Since both of these effects are likely to be stronger for extremists than for moderates, we should expect the coefficient on the squared DW-NOMINATE variable - which serves as a measure of extremism[35] to be positive. However, the sign of the base-level coefficient is less clear.

4.3 Modeling

To model patterns of institutional complexity within this dataset, I use a Bayesian gamma regression model with a hurdle component. We can loosely treat this model as a two-step model, in which we first estimate a logistic regression to determine whether a given observation is zero or non-zero and further estimate a gamma regression for the non-zero

[34] In particular, as Sinclair (2016) notes, since important and time-sensitive laws often bypass the ordinary lawmaking process the authors of these laws may not have the time to gather a substantial number of cosponsors. For an instructive example, contrast the Patient Protection and Affordable Care Act (Pub. L. 111-148) with the American Reinvestment and Recovery Act (Pub. L. 111-5). Though both laws received substantial public attention and the latter passed by larger margins than the former, the former law had 40 cosponsors while the latter had only 9. This difference likely reflects the speed with which the bailout law was enacted, compared with the more measured process for the Affordable Care Act.

[35] See, e.g., Theriault and Rohde (2011) for similar usage.

observations (since the dependent variable is continuous and strictly positive).[36] Within both components of the model, I estimate separate intercepts and coefficients for each of the eight predictor variables included in the model. For additional flexibility, I partially pool the intercepts by policy area in both the hurdle and gamma regression components of the model to capture policy area-specific effects.

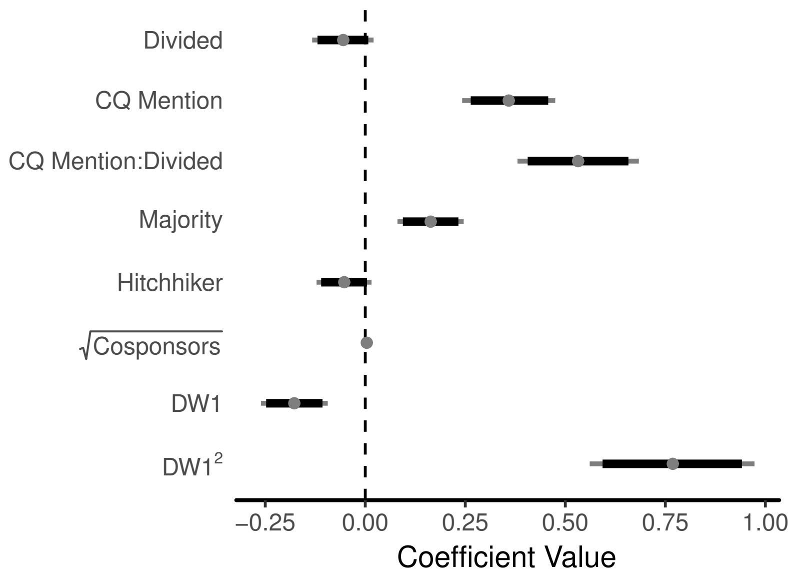
The motivation for this modeling structure follows directly from my theoretical expectations. The hurdle model I employ is an example of a mixture model, in which we treat the dependent variable as a mixture of two distinct latent "types". In the context of this project, I expect to encounter two kinds of laws: a "standard" type, which increases, decreases, or otherwise modifies the jurisdiction of one or more governmental actors, and a "technical" type, which does not alter the jurisdiction of any actor.[37] Laws of this kind are likely to follow a different data-generating process than other laws contained in my dataset. The hurdle component of the model separates these two law types, allowing me to estimate separate coefficients for each predictor and each type.

As I note in §4.2, one disadvantage to using the average degree of each law's implementing network as a dependent variable is that it leaves laws with zero named entities with an undefined value on the dependent variable. As a robustness check, I therefore follow Farhang and Yaver (2016) and calculate the *number of unique entities* named by each law as an alternate measure of institutional complexity. This measure is less conceptually precise, but is defined for all laws, which allows me to incorporate all observations into the models I present in §5. Like Farhang and Yaver (2016) I find that these two measures are highly related (correlation of $r = 0.91$ between log-plus-one transformed versions of each variable), and I draw essentially identical substantive conclusions using both sets of variables (see Appendix C.2.5 for details).

[36] See Appendix C.1 for details regarding model specification, priors, estimation, convergence diagnostics, and posterior predictive checks.

[37] E.g. Pub. L. 108-306, "To provide an additional temporary extension of programs under the Small Business Act and the Small Business Investment Act of 1958 through September 30, 2004, and for other purposes." This law simply extends authorization for existing provisions of the Small Business Investment Act of 1958, and therefore provides no modifications to existing administrative jurisdiction.

Figure 3: Top-level coefficient estimates, gamma regression model.



Grey dots indicate posterior mean values. Dependent variable is the average degree of each law's implementing network ($n = 4,296$). Thick lines indicate 90% credible intervals, and thin lines indicate 95% credible intervals. Positive estimates indicate that an increase in a given coefficient increases law complexity. Intercept suppressed for readability.

5 Results

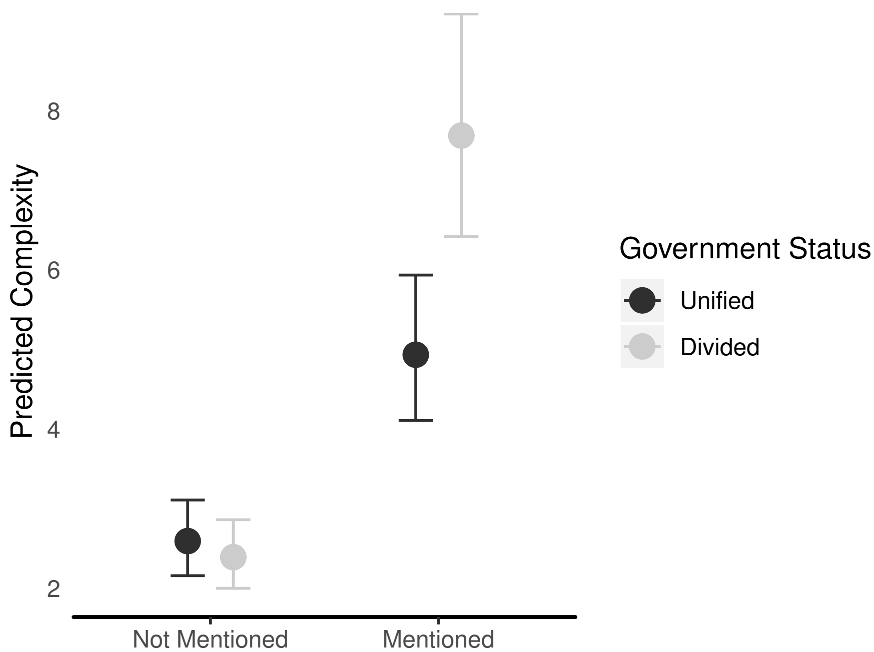
Top-level coefficients for the gamma regression model component - which models the complexity of "standard" laws that actually modify administrative authority - are given in Figure 3.[38] As predicted, the effect of divided/unified government is strongly conditioned by law importance. For non-"significant" laws - as measured by mention in CQ's year-end summary - the coefficient on the divided government indicator is small and possesses the opposite sign predicted in the literature. By contrast, "significant" laws passed under divided government are substantially more complex than their unified-government counterparts. Also as predicted, "significant" laws contain consistently more complex implementing structures than their "non-significant" counterparts, no matter the background political context.

These findings offer substantial support for the hypotheses I present in §2. Like other

[38]For brevity, I relegate coefficient estimates for the hurdle model - which estimates coefficients for "technical" bills - to Appendix C.2.1.

authors, I find that executive/legislative preference disagreements are associated with increased institutional complexity. However, this effect is almost entirely constrained to "important" laws that receive attention from legislators' constituents. As shown in Figure 4, for high-importance legislation the effect of executive/legislative preference disagreement is substantial: "important" laws passed under divided government are over 50% more complex than their unified-government counterparts. But, outside of this context, this difference essentially vanishes. This finding complicates findings presented elsewhere in the formal institutional design literature, and suggests that institutional design choices outside of high-importance issues are primarily driven by characteristics of the substantive policy problem under consideration.

Figure 4: Marginal effects plot, divided government/CQ mention interaction.



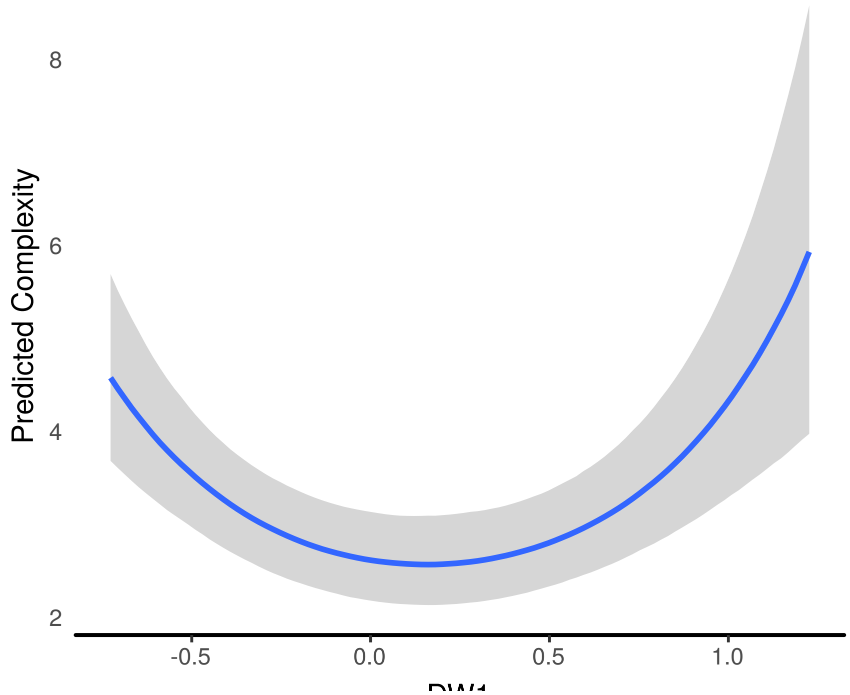
Margins plot, showing predicted values for the four conditions in the divided government/CQ mention interaction. Dependent variable is the average degree of each law ($n = 4,296$). Dots show posterior means and lines show 95% credible intervals. All other predictors in the model are fixed at zero.

Most of the secondary expectations I present in §4.3 are also supported. As predicted, members of the majority tend to pass laws containing more complex implementing structures, which suggests that these members tend to be willing to draft laws addressing more complicated policy problems. Similarly, "hitchhiker" laws tend to be less complex than independently-passed laws, though the coefficient estimate is small and the 95% posterior credible interval crosses zero. Perhaps more surprisingly, the coefficient on the consponsorship variable is essentially zero, which suggests that cosponsorship patterns are essentially unrelated to institutional design choices. However, as I show in the hurdle model results in Appendix C.2.1, laws that receive more cosponsors are more likely to be of the "standard" than the "technical" type. This finding fits with the existing cosponsorship literature; as authors like Wilson and Young (1997) and Box-Steffensmeier *et al.* (2019) suggest, cosponsorship is generally used by members as a position-taking tool early in the lawmaking cycle, and has a limited relationship with law content or downstream probability of passage. Since members likely derive little utility from taking positions on "technical" laws, we should likely expect these laws to attract few cosponsors.

The relationship between individual-level ideology and institutional complexity also largely matches my expectations. As I describe in §4.3, both liberals and conservatives possess incentives to propose laws containing complex institutional structures, and these incentives are likely to grow for more extreme members in both directions. As a result, it is not clear whether the base-level DW-NOMINATE coefficient should be positive or negative, but the DW-NOMINATE^2 coefficient - which is related to ideological extremism - should be positive. Figure 5 provides support for this set of expectations. For a given DW-NOMINATE magnitude, liberals (members with negative scores) tend to propose slightly more complex laws than conservatives, but the difference between centrists and extremists on either side is substantially larger than the left-right differential.[39] Future work should probe this finding further, and examine whether the relationship between ideology and institutional design remains constant across policy area and over a longer time horizon than the one I study in

[39] As I show in Appendix C.2.2, when the squared DW-NOMINATE variable is excluded from the model the corresponding coefficient estimate for the raw DW-NOMINATE variable drops to essentially zero. Substituting the DW-NOMINATE variable for an indicator denoting the party to which the proposing member belongs offers a similar conclusion, suggesting that this finding is robust to measurement and model specification choices. However, in all cases, coefficient estimates for other variables in the model are essentially unaffected.

Figure 5: Marginal effects plot, DW-NOMINATE score of proposing member.

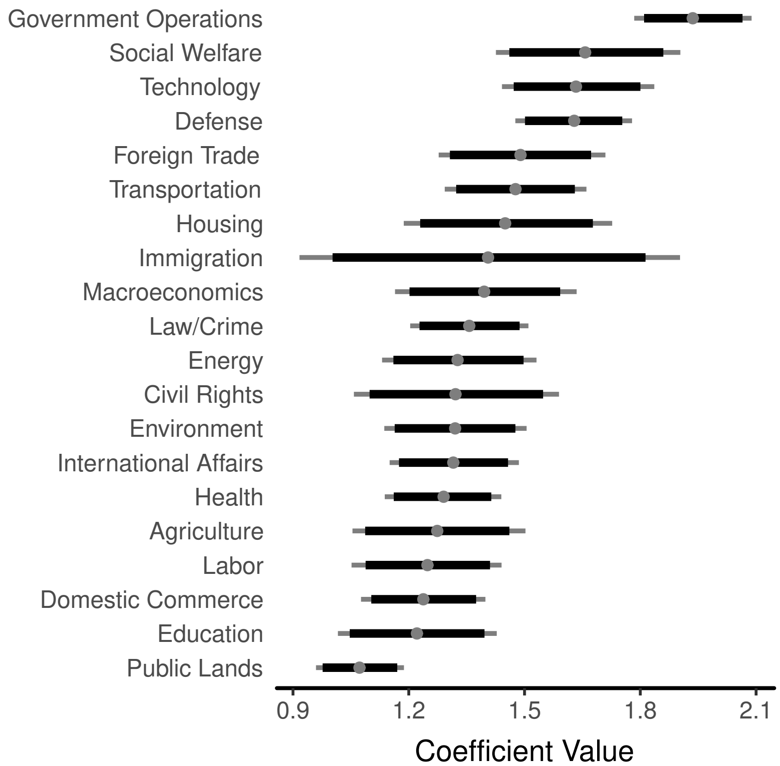


Margins plot, showing predicted values across various levels of the proposing member's DW-NOMINATE score. Dependent variable is the average degree of each law's implementing network ($n = 4,296$). All other predictors in the model were fixed at zero.

this paper.

Finally, the policy area predictions I offer in §4.3 are also at least partially supported. As shown in Figure 6, all else equal public lands laws tend to contain noticeably simpler implementing structures than essentially all other policy areas. This pattern is intuitive; public lands laws often involve small transfers of property between government institutions or small modifications in land regulation or oversight, rather than large-scale modifications of the administrative state. The top end of the scale also fits with existing knowledge, though findings are more mixed. Government operations and defense, for example, are classic high-complexity policy areas, which implicate a variety of different policy concerns and necessarily involve a large number of government agencies in their implementing process. However, outside of these relatively clear examples, policy area does not appear to have a strong effect on downstream institutional design decisions.

Figure 6: Intercept values by policy area, gamma regression model.



Intercept values, gamma regression model. Coefficient estimates are based on a partially pooled intercept term. Dependent variable is the average degree of each law's implementing network ($n = 4,296$).

6 Conclusion

Overall, this paper offers two primary contributions. From a theoretical standpoint, I argue that formal institutional design decisions are frequently unaffected by the background partisan context. In existing political science scholarship, institutional design choices - and particularly the *complexity* of formal institutions - are often treated as purely political tools, which legislators manipulate based on partisan circumstances. By contrast, I demonstrate that this relationship only holds for high-visibility, "important" legislation. Outside of this context, institutional complexity is primarily a function of the policy issues under consideration, rather than the partisan context. These findings dovetail with a growing literature on "submerged" bipartisan collaboration over the *content* of legislation (e.g. Wilkerson *et al.* 2015; Casas *et al.* 2018), and support a more optimistic portrayal of Congressional lawmaking than is often provided in the academic and popular presses.

To test this theory, I rely on a novel language- and network-based measurement strategy, which offers a straightforward, scalable way to extract information on formal institutional design choices. As I note in §2.3, extracting information on institutional design choices from legal texts is highly labor-intensive, which has prevented previous studies from moving beyond samples focused on single policy areas or "historically significant" legislation. By contrast, the method I present in this paper - which relies on both modern machine natural language processing methods and case-specific knowledge regarding Congressional drafting procedures - offers a more scalable alternative, which enables me to test the theoretical framework I present in this paper. Better still, the conceptualization and measurement strategy I employ is adaptable to other legal contexts, offering a useful tool for researchers interested formal institutional design patterns beyond the national-level American setting.

The results in this project offer a number of directions for future research. Within the domain of American legislation, comparing enacted legislation to bills that were proposed but not passed offers opportunities to understand the sources of institutional complexity in a direct fashion, both at an individual and an institutional level. Historical legislation offers other possible applications. For example, the approach I present in this paper might help researchers to identify whether the differences between laws enacted under divided and unified government remain constant as the ideological gap between the two parties grows and shrinks. More broadly, though this paper is restricted to American national-level legislation, the framework and tools I present are broadly applicable to legal documents from state governments or governments from outside the United States. As a result, the measurement approach I propose would help researchers to study institutional design patterns in different social and political contexts and under different systems of government. Finally, the measurement framework I present also offers opportunities to study other quantities of interest, such as the centrality of particular actors or the frequency with which pairs of actors collaborate on policy implementation.

References

Jason Anastasopoulos and Anthony M Bertelli. Understanding delegation in the european union through machine learning. 2017.

Isabelle Augenstein, Leon Derczynski, and Kalina Bontcheva. Generalisation in named entity recognition: A quantitative analysis. *Computer Speech & Language*, 44:61–83, 2017.

Frank R Baumgartner and Bryan D Jones. *The Politics of Information: Problem Definition and the course of public policy in America.* University of Chicago Press, 2015.

Frank R Baumgartner, Christian Breunig, Christoffer Green-Pedersen, Bryan D Jones, Peter B Mortensen, Michiel Nuytemans, and Stefaan Walgrave. Punctuated equilibrium in comparative perspective. *American Journal of Political Science*, 53(3):603–620, 2009.

Janet M Box-Steffensmeier, Dino P Christenson, and Alison W Craig. Cue-taking in congress: Interest group signals from dear colleague letters. *American Journal of Political Science*, 63(1):163–180, 2019.

Andreu Casas, Matthew Denny, and John Wilkerson. More effective than we thought: Accounting for legislative hitchhikers reveals a more inclusive and productive lawmaking process. 2018. Forthcoming: *American Journal of Political Science.*

Joshua D Clinton, Anthony Bertelli, Christian R Grose, David E Lewis, and David C Nixon. Separated powers in the united states: The ideology of agencies, presidents, and congress. *American Journal of Political Science*, 56(2):341–354, 2012.

John J Coleman. Unified government, divided government, and party responsiveness. *American Political Science Review*, 93(4):821–835, 1999.

Susan B Epstein and K Alan Kronstadt. Pakistan: Us foreign assistance. *Current Politics and Economics of the Middle East*, 4(3):575, 2013.

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

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

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

Fabio Franchino. Delegating powers in the european community. *British Journal of Political Science*, 34(02):269–293, 2004.

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

Roberto Franzosi, Gianluca De Fazio, and Stefania Vicari. Ways of measuring agency: an application of quantitative narrative analysis to lynchings in georgia (1875–1930). *Sociological Methodology*, 42(1):1–42, 2012.

William Howell, Scott Adler, Charles Cameron, and Charles Riemann. Divided government and the legislative productivity of congress, 1945-94. *Legislative Studies Quarterly*, pages 285–312, 2000.

Zhiheng Huang, Wei Xu, and Kai Yu. Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*, 2015.

John D Huber and Charles R Shipan. *Deliberate discretion? the institutional foundations of bureaucratic autonomy*. Cambridge University Press, New York, 2002.

John D Huber, Charles R Shipan, and Madelaine Pfahler. Legislatures and statutory control of bureaucracy. *American Journal of Political Science*, pages 330–345, 2001.

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

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. Neural architectures for named entity recognition. *arXiv preprint arXiv:1603.01360*, 2016.

Kalev Leetaru and Philip A Schrodt. Gdelt: Global data on events, location, and tone, 1979–2012. In *ISA annual convention*, volume 2, pages 1–49. Citeseer, 2013.

Xuezhe Ma and Eduard Hovy. End-to-end sequence labeling via bi-directional lstm-cnns-crf. *arXiv preprint arXiv:1603.01354*, 2016.

Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60, 2014.

Peter J May and Ashley E Jochim. Policy regime perspectives: Policies, politics, and governing. *Policy Studies Journal*, 41(3):426–452, 2013.

David R Mayhew. *Divided we govern.* Yale University, 1991.

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

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

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

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

Terry M Moe. Delegation, control, and the study of public bureaucracy. In *The Forum*, volume 10, 2012.

Jonathan Oberlander. The future of obamacare. *New England Journal of Medicine*, 367(23):2165–2167, 2012.

Barbara Plank, Anders Søgaard, and Yoav Goldberg. Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss. *arXiv preprint arXiv:1604.05529*, 2016.

Lance A Ramshaw and Mitchell P Marcus. Text chunking using transformation-based learning. In *Natural language processing using very large corpora*, pages 157–176. Springer, 1999.

Barbara Sinclair. *Unorthodox lawmaking: New legislative processes in the US Congress*. CQ Press, 2016.

Martin Sundermeyer, Ralf Schlüter, and Hermann Ney. Lstm neural networks for language modeling. In *Thirteenth Annual Conference of the International Speech Communication Association*, 2012. `https://www.isca-speech.org/archive/interspeech_2012/i12_0194.html`.

Sean M Theriault and David W Rohde. The gingrich senators and party polarization in the us senate. *The Journal of Politics*, 73(4):1011–1024, 2011.

Michael M Ting. A strategic theory of bureaucratic redundancy. *American Journal of Political Science*, 47(2):274–292, 2003.

Rachel VanSickle-Ward. *The Devil Is in the Details: Understanding the Causes of Policy Specificity and Ambiguity*. SUNY Press, 2014.

Craig Volden and Alan E Wiseman. *Legislative Effectiveness in the United States Congress: The Lawmakers*. Cambridge University Press, 2014.

John Wilkerson, David Smith, and Nicholas Stramp. Tracing the flow of policy ideas in legislatures: A text reuse approach. *American Journal of Political Science*, 59(4):943–956, 2015.

Rick K Wilson and Cheryl D Young. Cosponsorship in the us congress. *Legislative Studies Quarterly*, pages 25–43, 1997.

James Q Wilson. Bureaucracy: What government agencies do and why they do it. 1989.