Cognitive Load and Issue Engagement in Congressional Discourse

Robert Shaffer[1]

University of Texas, Austin

Abstract

In ordinary conversation, cognitive limitations prevent speakers from engaging with high-dimensional policy problems. However, in certain situations, individuals may be willing to shoulder the larger "cognitive load" associated with these kinds of issues. For example, in the US Congress, members in a leadership role and members grappling with so-called "focusing events" may be particularly inclined to engage with a broad range of topics compared with their backbench and lower-engagement colleagues.

Existing research has documented this expansionary pattern at the institutional level, demonstrating changes in issue dimensionality in the US Congress in response to to various external stimuli. However, data and measurement limitations have prevented these studies from addressing individual-level questions. In this paper, I attempt to fill this gap using an approach drawn from the computational text analysis literature. As I argue, computer-assisted content analysis tools are well-suited for detecting changes in issue diversity within a corpus of texts. I demonstrate the potential utility of these approaches by examining patterns in issue dimensionality in Congressional hearing transcripts, with a focus on hearings surrounding the 2008-2009 Financial Crisis.

1. Introduction

In decision-making scenarios, a key challenge relevant actors face is the problem of managing *issue dimensionality*. When deciding on consequential matters, actors frequently grapple with a dizzying array of information and problem dimensions. For a concrete example, consider national-level economic policy. Even straightforward changes to macroeconomic regulations (e.g. capital requirements for banks) force Congress to address a wide variety of downstream

Email address: rbshaffer@utexas.edu (Robert Shaffer)

effects, including inflation, unemployment, business debt pricing, and mortgage availability.

Usually, cognitive limitations prevent individuals from considering all aspects of a particular issue [20, 38]. Political institutions are guided by this same general cognitive phenomenon. Like the individuals that compose it, Congress can only focus on a few ideas at a time, leaving most problems to languish until a crisis point is reached. As a result, aggregate-level policymaking is often characterized by a so-called "stick-slip" pattern, in which framing and high-level policy remain stable for long periods of time interrupted by "punctuations" of increased interest and activity [e.g. 22, 2].

In political settings, these patterns of information ingestion have important implications for institutional efficacy. Popular and academic writers alike often characterize Congress as dysfunctional or even "broken" [e.g. 26]. As Adler and Wilkerson [1] argue, this criticism is incomplete at best. Though Congress often experiences intense disagreement on high-visibility issues, at least on so-called "compulsory" policy problems (i.e., those issues that are pressing enough to compel a Congressional response), legislators of different parties are often able sift through a variety of problem dimensions and deliver a legislative solution.

However, beyond these aggregate-level findings, we know relatively little about the individual-level factors driving Congressional policy discussions. For example, how long do periods of high issue diversity last? Which individuals are most willing to raise a broad set of problem aspects? And, are particular types of actors (e.g. leadership members) more willing to raise a broad set of issue dimensions than others? The interplay between these strategic considerations and individual cognitive limitations are important from an academic standpoint, but are also relevant for broader efforts to assess Congressional efficacy. By identifying the situations and types of actors most inclined to consider a diverse set of problem aspects, activists and academics alike can more easily identify the individuals and institutional arrangements most likely to produce productive policy discourse.

In this paper, I address these questions through a measurement approach drawn from computational linguistics. At least in the political domain, individual-level text data are becoming broadly available, as policy-making bodies like the US Congress digitize and release transcripts of important policy conversations (e.g. committee hearings or speeches). As I argue, topic models and other unsupervised dimensionality reduction tools are well-suited for detecting changes in issue attention and diversity of conversation across large collections of text. I outline one approach based on these methods, and apply this approach to examine patterns in conversation dimensionality in the US Congress from 2004 to 2011. I find that of Congressional discourse spiked among all subgroups by approximately 15% around the onset of the 2008-2009 Financial Crisis. Moreover, I find that dimensionality varied in predictable ways throughout the time series, with incentivized speakers (e.g. leadership members) engaging more deeply with relevant policy problems than their less-engaged counterparts.

2. Issue Engagement and "Cognitive Load"

2.1. The Politics of Problem Dimensionality

In the broader decision-making literature, an important theme for many studies is the notion of *cognitive constraints*. As Simon [1985, 1996] argues, individual behavior in decision-making settings can be best described as intendedly rational. Though human actors often pursue goal-directed, utility-maximizing patterns of behavior, their ability to achieve these goals is constrained. In particular, individuals possess limited ability to consider and compare the relevant dimensions of various problems, creating a sort of "oversupply of information" [47] that decision-makers cannot easily process. As a consequence, when faced with high-dimensional problems, individuals resort to cognitive shortcuts, processing problem dimensions serially and relying on third-party signals and other decision-making heuristics [e.g. 45, 27, 20].

The difficulties involved with the ingestion of new information can be usefully framed through the concept of "cognitive load." As defined in the instructional design and problem-solving literatures, the "intrinsic cognitive load" of a particular task refers to "demands on working memory capacity [...] intrinsic to the material being learned" [35]. Some tasks (e.g. elementary algebra and numerical reasoning) are relatively simple and require little effort to absorb, while others (e.g. calculus and higher-level mathematics) require substantially greater time and attention to master [41]. Though the cognitive load of a particular task is usually presented as an immutable aspect of that task, instructors (or other actors with agenda control) can break concepts into simpler "chunks" [11] or eliminate non-germane problem aspects [42] in order to ease individual-level cognitive demands.

Translated to the political domain, these ideas are complicated by the presence of strategic behavior. Consider committee activity in the US Congress. When such efforts are politically useful, committee leaders frequently attempt to claim ownership over particular problem areas [23], or to articulate and promote a favorable understanding of certain problem domains [43, 3]. These moving jurisdictional boundaries create difficulties when researchers attempt to categorize policy problems into appropriate "chunks" whose complexity can be easily measured. The 2007-2008 Financial Crisis provides a concrete example of this phenomenon. Though financial policy might seem like the domain of committees like the Senate Finance Committee or House Financial Services Committee, the crisis also attracted attention from the House Committee on Agriculture and the Senate Committee on Homeland Security and Government Affairs, who sought to "take ownership" over particular aspects of the crisis that were relevant to their respective portfolios.

The interplay between strategic considerations and cognitive constraints extends beyond issue definition, and into other areas of political behavior. As McCubbins and Schwartz [30] famously argue, Congressional oversight and lawmaking behavior can be (loosely) categorized into two conceptual categories, which they term "police patrol" and "fire alarm"-style activity. In the former case, legislators regularly "patrol" bureaucratic activity, issuing closely-written legislative directives and maintaining constant oversight over a shifting sample of administrative actions. By contrast, under the crisis-based "fire alarm" model, legislators let oversight activity in particular areas languish for long periods until third-party actors (usually, citizens or interest groups) draw attention to particular problems. McCubbins et al. present this behavioral pattern in a classic rational-choice framework, and argue that "fire alarm"-type oversight behavior represents a rational allocation of limited cognitive and financial resources:

> When legislators try to write laws with sufficient detail and precision to preclude administrative discretion, they quickly run up against their own cognitive limits: beyond a certain point, human beings just cannot anticipate all the contingencies that might arise. The attempt to legislate for all contingencies can entail unintended (and undesired) consequences [30, 175].

By empowering citizen groups to conduct oversight action on their behalf, legislators can exploit bureaucratic expertise without attempting to prespecify limitations on administrative power. This approach is both cognitively and strategically appealing. By delegating oversight authority and establishing broad "framework"-style legislation, lawmakers can set policy in less salient areas while engaging with a narrow set of issue dimensions, leaving them free to dive more deeply into pressing problems [see also, e.g. 25, 29, 31].

These basic insights have inspired a massive quantity of scholarship in the political science and public policy domains, which I can only briefly discuss here. In general, researchers in this area are primarily interested in understanding the circumstances under which decision-makers are more willing to devote attention and cognitive energy to a particular policy area, and the mechanisms by which individuals attempt to overcome cognitive constraints [20]. From the theoretical side, the role of signaling and private information represents a core area of interest for many researchers throughout the social sciences. In the literature on executive/legislative dynamics, in particular, researchers have attempted to model a variety of interactions between nominally-subordinate experts (e.g. committees and executive agencies) and elite decision-makers (e.g. the Congressional floor, the President) [15, 5, 13, 14].

Studies in this area are also frequently concerned with the rules institutions create to manage their own behavior, particularly in the context of administrative governance and institutional control [34, 31, 33]. As groups of individuals, organizations and institutions like the US Congress are constrained in their ability to ingest and process new information, and frequently create routinized, automatic processes which are slow to adjust to changes in external conditions [2, 21, 22]. Work associated with the Comparative Agendas Project (CAP) represents perhaps the best-known empirical body of work in this area; using a comprehensive dataset of newspaper articles, legislation, committee hearings, and political speeches from the US and around the world, researchers involved with the project have analyzed variation in budgeting and allocation of attention to various policy issues on offer, and found substantial evidence of the kind of "stick-slip" patterns predicted in the bounded rationality literature [4].

2.2. Shouldering the Load: Individual-Level Predictions

Despite the volume of work in this area, an array of important questions remain unanswered. In particular, few existing empirical studies in this literature actually measure and study individual-level behavior. Major datasets like CAP primarily code aggregated, institution-level outputs (e.g. Congressional bills or committee hearings), rather than individual-level observations (e.g. individual member amendments to bills, or statements within committee hearings). Because of these limitations, datasets like those generated by CAP cannot easily study relationships between individual-level attributes and attention distribution phenomena.

To highlight the gaps left by existing approaches, I focus on three basic hypotheses regarding the interplay between individual-level cognitive limitations and strategic considerations. Though individuals are generally disinclined to assume tasks with a heavy cognitive load, there are at least some situations in we might expect individuals to engage with a more diverse set of issue dimensions:

2.2.1. Crisis Events

During "fire alarm"-type crisis events or other periods of intense interest, lawmakers may be more willing than usual to devote attention to a given problem area, and to accept the accompanying cognitive costs. The financial crisis provides an acute example of a crisis event, which should provoke individuals to ingest an especially broad quantity of information relative to their previous patterns of issue engagement.

2.2.2. Expertise

As noted previously, the cognitive load imposed by a particular task depends on individual familiarity with the subject area at hand. As a result, we should expect expert witnesses to be more willing to engage deeply with the subject matter of a particular hearing compared with non-expert members of Congress. This general trend should be particularly noticeable in witnesses with broad expertise and experience in a given policy area (e.g. high-level career bureaucrats with repeat experience giving Congressional testimony).

Norms of committee discourse reinforce this expectation. Since members of Congress are allowed to question witnesses on topics of their choosing, most witnesses will be compelled to testify on a broad array of topics. Members, by contrast, can restrict their discussion to topics of their choice.

2.2.3. Leadership

In Congressional committees, individuals with a stake in the size of the committee's jurisdiction (e.g. members in a leadership role) may be more willing to adopt and press a more expansive (and more cognitively taxing) view of a given committee's agenda. As a result, these members are likely to speak on a broader range of topics than their back-bench counterparts. Again, the organization of Congressional hearings reinforce this expectation. Leadership members frequently make broad opening and closing statements in each hearing, which set the agenda and the general themes for the day's discussion. These statements provide an additional opportunity for leadership members to speak in a wide-ranging fashion, allowing them to discuss a broader range of ideas.

Not all variation in conversation dimensionality is driven by cognitive effects. Institutional rules likely matter as well; for example, leadership members of Congress are given more opportunities and more freedom to speak about a broad range of topics, while witnesses are often compelled to speak broadly no matter their particular areas of expertise. As a result, these cognitive and institutional effects are likely not separable. At the very least, however, any changes observed in response to environmental factors should be directly attributable to cognitive responses to crisis events.

3. Methodology

3.1. Text Analysis and Automated Dimension Extraction

As mentioned previously, little existing work in the information-processing and agenda-setting literatures focuses specifically on individual-level decision-making. Largely, this limitation results from a measurement problem. Though researchers have developed large datasets on organizational priorities (e.g. the Comparative Agendas Project's coding efforts on news stories, Congressional bills, and committee hearings), no comprehensive dataset of individual-level distribution of attention exists, even for elected officials and other policy elites.

Text analysis offers an alternative approach. In recent decades, governments and other organizations have become increasingly willing to release conversation transcripts, hearings, speeches, legislative documents, and similar individual-level data sources. These texts (potentially) reveal important information about individual-level preferences, offering new ways to investigate the questions and hypotheses given above.

In order to measure dimensionality and issue engagement in text data, we need new modeling and measurement strategies. In particular, we need a consistent approach by which to identify and categorize the topics under discussion at a particular moment in time. In other settings, researchers have relied on intensive hand-coding procedures to code documents in this fashion, employing research assistants to read and categorize documents of interest (for example, in the Comparative Agendas Project). Unfortunately, this approach is difficult to implement when individual behavior is the phenomenon of interest. Ideally, to study individual-level issue engagement we would need to code each statement made by each individual according to the proportion of time spent on each topic of interest, producing a compositional vector $p = \begin{bmatrix} e_1 & e_2 & e_3 & ... & e_n \end{bmatrix}$ with n topics, $e_i \geq 0 \forall i \in \{1, 2, ...n\}$, and $\sum_{i=1}^{n} e_i = 1$.

Unfortunately, coding committee transcripts and other large data sources by hand is essentially unworkable. Large documents with many speakers and many topics of interest are too labor-intensive for human readers to parse at the necessary speed, forcing human-coded datasets to restrict themselves to coding higher-level organizational priorities. For example, for practical reasons the Comparative Agendas project codes committee hearings based on the overarching theme of a given hearing, rather than coding individual statements or the topical composition of the hearing as a whole. As a result, studies that rely on these kinds of data sources are generally restricted to asking institutional-level questions, which miss important individual-level phenomena of interest.

Machine-assisted methods, by contrast, are more promising. Over the last several decades, as computing power has become cheaper, computer scientists and statisticians have become increasingly interested in using statistical methods to analyze unstructured text data. A large body of work in this area has focused on unsupervised extraction of latent themes or ideas from textual corpora, which can be used to summarize the content of large, multidimensional documents. Prominent approaches in this area include latent Dirichlet allocation (LDA) and its many variants and extensions [e.g. 9, 7, 8, 16, 37], and more recent deep-learning based approaches such as Mikolov et al. [32]'s word2vec. Estimation and modeling details vary across these modeling approaches; however, all essentially attempt to reduce high-dimensional textual data into some lower-dimensional representation. LDA and its variants, in particular, naturally estimate the same compositional proportion vector described above for each document and each extracted topic, making that modeling approach a natural choice for this kind of analysis.

No data-collection or modeling approach is perfect, and the tools described above are no exception. In particular, as noted earlier in this section, the dimensions (or "topics") extracted by these models are (usually) produced in an unsupervised fashion. Users therefore have no guarantee that the topics produced by these models will map onto preexisting conceptual schemes. However, in other applications, researchers have found that models trained on policy-related corpora (e.g. news coverage or Congressional speeches) tend to produce topics that are related to human conceptualizations of the policy space [see, e.g, 24]. Application-specific validation of extracted topics is necessary in order to be certain that the extracted topics are related to the substantive phenomena of interest [17]; however, these approaches represent promising starting points for the analysis of attention dynamics in large text corpora.

3.2. Dimensionality in Text: A Measurement Strategy

After obtaining a compositional topic proportion vector like the one described above, we need some way to convert that vector into a measure of dimensionality. In the human-coding scenario, a natural approach might be to generate a conceptual "inventory" of all topics present in the dataset (updated as new topics are introduced), and calculate a dimensionality statistic that reflects the degree that particular document "focuses" on a particular topic or set of topics. Unfortunately, in the machine-coding scenario, most modeling approaches can only extract a fixed number of topics from a particular corpus [though see 44], making the human-coded scenario described above difficult to implement.

As a result, in research settings where variation in the number of topics (dimensions) is the object of interest, a slightly different approach is more common. Given a fixed number of topics, models like LDA essentially attempt to simultaneously identify the latent topics and estimate the proportion of each document devoted to each latent topic. For any particular corpus, researchers can therefore estimate a model with the number of bins n in the topic proportion vector p as a fixed value. Then, given that "inventory" of topics, they can calculate the degree to which a particular document is "focused" on one or a few of the available topics. This approach requires researchers to re-estimate a model for each distinct corpus or application, and makes it difficult to compare results from various models to one another. However, assuming that the extracted dimensions do map onto conceptual categories of interest, this modeling approach produces an internally consistent dimensionality measure, allowing researchers to compare individuals or groups represented within the corpus of interest.

In selecting a dimensionality statistic, I adopt and extend the approach suggested in Boydstun et al. [10]. In particular, I use the so-called "effective topics" statistic as a measure of dimensionality:

$$\tau = n^{\eta}$$

Where n is the number of bins in the topic proportion vector p and η is the normalized informational entropy of p, defined as:

$$\eta = \frac{1}{log(n)} \sum_{i=1}^{n} p_i log(\frac{1}{p_i})$$

Informational entropy is a measure of concentration in a compositional vector. An informational entropy of $\eta = 0$ indicates that all verbiage in the given vector p is devoted to a single topic, while an informational entropy of $\eta = 1$ indicates that the vector p splits its verbiage evenly across all topic values.

As Boydstun et al. [10] show, the informational entropy statistic is highly nonlinear, making interpretation difficult. To aid interpretation, I therefore define the *effective topics* transformation:

$$\tau = n^{\eta}$$

For any η, τ can be interpreted as the number of equiprobable bins required to produce the given entropy value η.[1] For example, for $n = 40$ topics, if a particular person divided their verbiage between those categories such that $\eta = 0.5$, a second person who divided their verbiage evenly between $\tau = 40^{0.5} \approx 6.32$ topics would also produce an equivalent entropy value $\eta = 0.5$.

For the remainder of this paper, I present results on the linearized effective topics scale rather than the non-linear informational entropy scale. This transformation offers a more straightforward interpretation than the usual approach, and allows readers to compare differences in a linearized fashion.

3.3. Summary

Overall, then, the measurement strategy proposed in this paper proceeds as follows:

1. **Identify** a corpus of documents addressing the population group of interest.
2. Using that corpus, **estimate** a dimensionality-reduction model (e.g. LDA) and validate its results.
3. **Calculate** dimensionality statistics (e.g. entropy or the effective-topics measure), and use those statistics to compare dimensionality patterns across the corpus.

These steps each gloss over substantial complication and validation steps, some of which I highlight in the applied section of this paper. However, the basic strategy is fairly straightforward. Modern model-fitting and data cleanup tools make the estimation step relatively simple, and provide easy-to-use model validation tools. Once the model has been generated and validated, we can use statistics like the effective-topics formula to conduct comparisons of interest, and answer important questions about issue prioritization and engagement that were previously difficult to address.

4. Data

4.1. Issue Dimensionality in Congressional Discourse

As an application of these techniques, I examine dimensionality patterns in an original dataset of Congressional hearing transcripts. This dataset consists of

[1] See Appendix A for proof and discussion

all hearings posted on the Government Publishing Office (GPO)'s website from 2004-2011, parsed by speaker and linked with metadata drawn from the GPO's website and from Stewart and Woon [40]'s committee membership data via a series of custom Python tools (described in greater detail in Appendix B).

This dataset consists of all hearings posted on the Government Publishing Office (GPO)'s website which were coded by the Comparative Agendas Project as related to Macroeconomics, Community Development and Housing, or Banking, Finance, and Domestic Commerce. I then split these hearings by statement and linked each statement with speaker-level metadata (e.g. party, speaker type, speaker seniority) drawn from the GPO's website and from Stewart and Woon [40]'s committee membership data.[2] These hearings were selected based on their relationship to known policy areas affected by the financial crisis [see, e.g., 19], and represent both a substantive important subset of Congressional discourse and a useful validation set for the methods described in this paper. The GPO's database begins in approximately 2004 and ends in 2016; however, since Stewart and Woon [40]'s committee membership data is not available beyond the 112th Congress, the time period for this dataset is restricted to 2004-2011.

Fortunately, the 2004-2011 time period is of particular interest because it allows us to conduct a within-time series comparison of issue engagement patterns before and after the 2008-2009 Financial Crisis. Anecdotally, following the crisis members of Congress appeared to broaden policy discussions to cover an array of issues mentioned as potential causes for the crisis, as well as accompanying policy tools that might be used to prevent future crisis events. Example topics of discussion included swaps and derivatives, mortgage-backed securities, foreclosure rules, liquidity requirements, and the impacts of these devices and regulations on both the business environment and individual constituents. However, with the measurement strategy outlined above, we can study Congressional discourse more systematically, allowing us to test and validate these anecdotal observations.

4.2. Dataset Creation and Model Fitting

To convert the hearing data described above to a machine-interpretable dataset, I transformed the dataset to a bag-of-words representation. In this setup, each document is converted into a word-count vector, consisting of a series of a count of the number of times each unique term in the datset occurs in each document. This representation discards word order but retains document-level word covariance information, which forms the basis for most text analysis models. Next, I conducted a series of cleanup steps (described in detail in Table 1). Informally,

[2] Since the hearing transcripts contained in these data do not contain embedded information, speakers had to be matched to appropriate metadata using a heuristic-based process described in Appendix B). Approximately 87% of statements were successfully matched using this procedure.

Table 1: Pre-processing specification.

Terms	Documents	Other
Terms $\leq$ 3 characters, terms occurring in $\leq$ 10 documents discarded	documents $\leq$ 5 words discarded	lower-case, punctuation discarded, stopwords[a] discarded

[a] Stopword list drawn from NLTK's stopword corpus.

these preprocessing steps serve to discard short documents and rare and common terms (e.g. modifiers and articles), as well as to map certain term variants (e.g. upper/lower-case terms) to a common base.

In the text analysis context, discarding rare and common words serves two purposes. First, from an analytical perspective, very rare and very common words are not likely to be informative. Words that only occur in one or a few documents in a given dataset are frequently either mistyped or are specific to a small proportion of the dataset, and are not substantively relevant to the larger analytic task. Similarly, common words like "and" and "the" carry little substantive information about matters of interest, and can usually be safely discarded.

Second, very rare and very common words are computationally difficult to manage. As mentioned earlier, most text analysis models rely on document-level word covariance to learn about underlying model parameters. By definition, rare words co-occur with few other terms, which leaves most statistical tools with relatively little information to harness. In most cases, then, dropping words that occur in a few documents sacrifices little analytical leverage while easing computational burden.[3] Similarly, common words co-occur with many terms in the dataset, making it difficult for computational models to distinguish between them.

After completing these cleanup steps, the remaining dataset contained approximately 98,000 statements drawn from 582 hearings and 23 distinct committees. Using this cleaned dataset, I then fit a 40-topic latent Dirichlet allocation (LDA).[4] In its most basic form, LDA is a three-level Dirichlet-Multinomial hierarchical model, which models documents as mixtures of latent "topics" (prob-

[3] In the broader literature, dropping larger sets of terms (e.g. terms that occur in fewer than 1% of all documents) is common [17]. As Denny and Spirling [12] note, this preprocessing choice can discard important information, and can reduce model performance. Here, however, I only discard words that occur in ten or fewer documents ($\sim 0.01\%$ of the dataset), a much lower cutoff than is usually used. As mentioned in-text, this cutoff is intended to discard mistyped terms and terms that are specific to a very small subset of documents, and is unlikely to incur the kinds of issues found elsewhere in the literature.

[4] With the model fit via a Gensim wrapper for MALLET [36, 28] and the asymmetric prior setup described in Wallach et al. [46]. For a more detailed description of the LDA estimation and parameterization steps, see e.g., Blei [6].

ability distributions over words). Loosely, the generative process contained in LDA can be described as follows. Suppose each document in the corpus of interest consists of a vector of word counts, with the length of the vector equal to the total number of unique words in the corpus. Given a fixed number of topics K, documents are constructed as follows [6]:

1. Conditional on the observed word counts, draw a distribution over latent "topics."
2. For each topic, draw a distribution over words (i.e. a probability mass function that describes the probability of drawing each word conditional on being in the given topic).
3. Conditional on the topics constructed in (1) and (2), for each word in the document:
   (a) Draw a topic from the distribution over topics in (1).
   (b) Draw a word from the topic distribution constructed in (2).

As noted earlier, the choice of the dimensionality parameter K is application-specific, and depends on researcher judgment. For the dataset described in this paper, I fit a set of models with $K \in \{20, 25, 30, ...100\}$, and inspected top-probability words in each fit model. Based on these results, a 40-topic model seemed to offer a reasonable balance between topic coherence and excessive granularity, which I present in all subsequent analyses in this paper. At least in this application, however, the choice of the dimensionality parameter does not appear to affect substantive results (see Appendix D for details).

Next, I inspected the results from the 40-topic model described above, and assigned labels to each topic.[5] Like most topic modeling applications, this setup produced a mix of conceptually useful and "junk" topics, which do not map onto any conceptual category of interest. Examples of each type are given in Figure 1.[6]. For robustness, I experimented with dropping "junk" topics from the dataset,[7]. However, inclusion or disinclusion of these "junk" topics did not appear to affect the results given in the following section.[8]

Finally, after cleaning the dataset and inspecting model results, I summed statement-level topic proportion vectors into speaker-hearing combinations (weighted by word count of each statement), normalized each summed vector, and calculated an effective-topics value for each combined observation ($n \approx 10,000$). With this procedure, comparability of topics is a potential concern. Statistics like entropy (and the derived effective-topics transformation) implicitly assume that

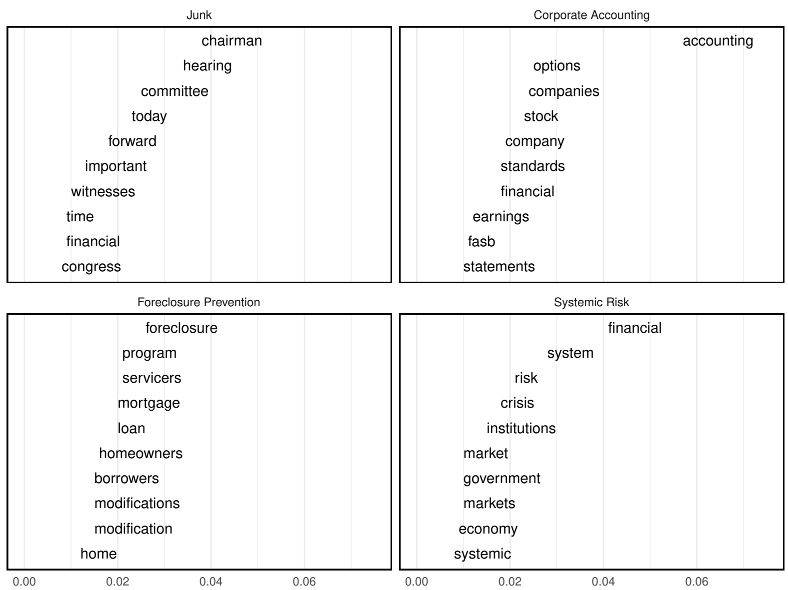
[5] Specifically, two coders independently read the 20 highest-probability words from each topic and the ten documents (statements) from each topic that contained the highest proportion of their words drawn from each topic. Then, each coder assigned labels to each topic and reconciled their results, giving final topic labels.

[6] See Appendix C for all top-probability words in the 40-topic model

[7] To drop "junk" topics, I discarded all words identified by the model as belonging to a "junk" topic, and normalized the remaining bins to form an updated topic proportion vector.

[8] See Appendix D for details.

Figure 1: Top-probability words for selected topics. Words are positioned such that the leftmost edge of the word indicates the probability of drawing that word for a given topic.



all topics cover a similar substantive scope. Since unsupervised dimensionality reduction tools like LDA are not guaranteed to return substantively comparable topics, we might be concerned that some topics returned by the model cover a narrower range of issues than others. While this concern is difficult to address directly, if varying topical scope was a concern we would expect models with different numbers of topics to return different substantive results (since larger models would likely "subdivide" certain issue areas further than their smaller counterparts). Thankfully, as mentioned earlier, fitting models with varying numbers of topics does not appear to affect the results given in this paper.[9]

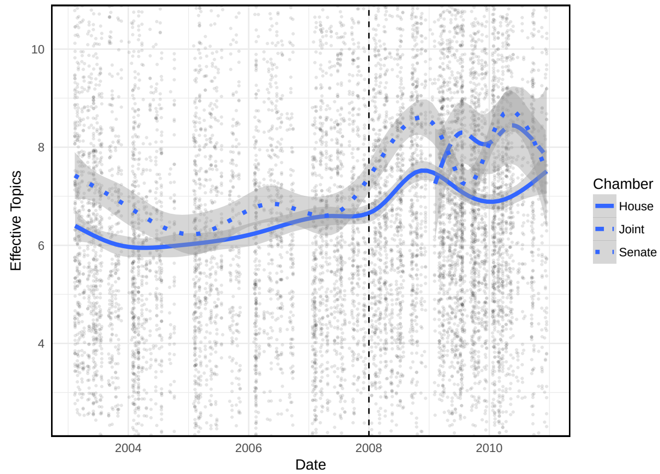
5. Results and Discussion

5.1. Aggregate-Level Issue Dimensionality

One primary motivation for building the financial policy corpus described above is to examine the impact of the financial crisis on financial policy discourse. Defining the start of the crisis as January 1, 2008 (the date at which the US economy officially entered a recession), conversation dimensionality in financial

[9]See Appendix D for details.

Figure 2: Effective topics values for combined speaker-hearing proportion vectors, separated by the chamber in which the hearing was held. Dashed line indicates January 1, 2008 (the official start of the Great Recession).

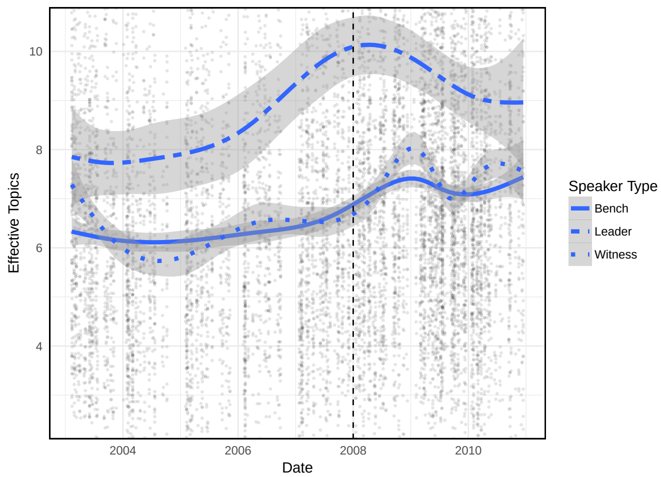


policy-related hearings expanded by approximately 15% post-crisis (7.3 effective topics versus 6.4; $p < 10^{-16}$, Welch's t-test). As shown in Figure 2, the pre/post-crisis difference in conversation dimensionality was present in both the House and the Senate, with both chambers experiencing a sharp spike in conversation dimensionality around the start of the crisis and leveling off as the crisis subsided. However, for most of the post-crisis time period covered by this dataset, the Senate displayed a larger shift than the House, as well as a higher average effective topics value overall (7.4 effective topics versus 6.7, $p < 10^{-16}$, Welch's t-test).

We can use this same approach to examine aggregate patterns in dimensionality by speaker type. Throughout the time period covered by the dataset, leadership members exhibit consistently higher effective topics values than witnesses and their backbench counteraprts (9.0 effective topics versus 6.8; $p < 10^{-16}$, Welch's t-test). All subgroups, however, experience a similar post-crisis spike and subsequent decline in conversation dimensionality, suggesting that the basic cognitive impact of the crisis event remains constant across all segments of the dataset.

This basic pattern seems reasonable, and fits with our existing understanding of the interplay between cognitive constraints and strategic behavior. Like many high-level policy problems, financial policy is complex and multidimensional, encouraging members to turn to heuristic devices (e.g. party platforms and expert recommendations) to avoid assuming a heavy cognitive burden. During crisis periods, traditional heuristic devices break down, forcing individuals to

Figure 3: Effective topics values for combined speaker-hearing proportion vectors, separated by speaker type. “Leadership” members are defined as those members holding a committee chairmanship. Dashed line indicates January 1, 2008 (the official start of the Great Recession).



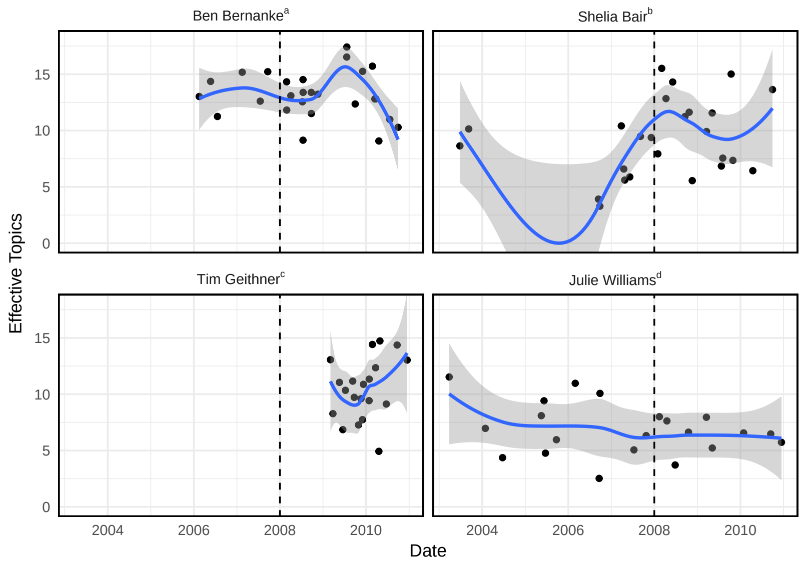
explore problem aspects and approaches outside of preexisting paradigms [20]. However, the incentive to dive more deeply into a particular problem area is likely not constant across all policymakers and members of Congress. Members of the Senate, for example, usually possess larger and more diverse constituencies than their House counterparts, and possess a greater incentive to address diffuse policy problems such as macroeconomic events. Leadership members possess similar incentives to speak broadly, as well as institutionally-defined opportunities to speak on a wide range of topics (e.g. agenda control through opening and closing statements).

The pattern of decline following the end of the crisis also fits with this explanation. As the crisis begins to recede in institutional memory, members should begin to adopt new problem understandings and heuristic devices, reducing the scope of the debate. Surprisingly, at least in the time period covered by this dataset, conversation dimensionality does not appear to have returned to its pre-crisis level of diversity, and may have actually increased slightly towards the end of the time period examined. Tracking the duration and size of post-crisis event changes in conversation dimensionality across other policy areas and time periods represents a promising direction for future work.

5.2. Individual-Level Attention Patterns

Besides these high-level institutional comparisons, we can also examine the behavior of particular individuals. Due to data availability problems, comparisons

Figure 4: Effective topics values for selected witnesses.



[a]Chairman of the Federal Reserve (2006-2014).
[b]Professor of Financial Regulatory Policy at the University of Massachusetts (2001-2006) and Chairwoman of the Federal Deposit Insurance Corporation (2006-2011).
[c]Secretary of the Treasury (2009-2013)
[d]Deputy Comptroller of the Currency (1994-2012)

between individuals are inherently more challenging than higher-level institutional comparisons. Since I aggregate individual statements into speaker-hearing combinations, even the most verbose individuals only speak at a few hundred hearings, leaving a relatively small set of data points to examine. Fortunately, though, many of those same verbose individuals are also the most influential, making them useful case studies for further analysis on individual-level patterns in conversation diversity.

As an example of this type of analysis, consider patterns in witness attention dynamics. Figure 4 shows the effective topics values for each speaker-hearing time for four of the most speakers in the corpus, and demonstrates some basic patterns in hearing organization and discourse. Take, for example, then-Federal Reserve Chairman Ben Bernanke. Compared to other common witnesses, Bernanke addresses a noticeably larger array of topics throughout the time period under consideration. This pattern fits with our knowledge about the particular relationship between Congress and the Federal Reserve. Even in non-crisis periods, the Chairman of the Federal Reserve is legally required to testify before Congress on a semiannual basis, and report on the state of the economy. These reports are remarkably wide-ranging; in Bernanke's first Congressional report, for example, he addressed standard monetary policy concerns such as GDP growth, unemployment, and inflation, but also income inequality, solvency of entitlement programs (e.g. Social Security and Medicare), and the impact of Hurricane Katrina on global energy markets and supplies.[10] Based on these legal requirements alone, then, we should expect Bernanke to cover an unusually diverse array of topics in his testimony compared with other witnesses in the dataset.

Other witnesses, by contrast, are much more focused. Unlike the Federal Reserve Chairman, who is explicitly called upon to regularly express his or her views on a very wide variety of topics, other witnesses are generally asked to comment on specific bills or policy problems, creating a more mixed record. Take then-Secretary of the Treasury Timothy Geithner. During the period of study used in this paper, Geithner was called upon to testify on foreclosure reduction and assistance programs[11], the bailout of AIG[12] - both relatively focused topics - as well as broader ideas such as derivatives and their impact on systemic risk[13]. The latter hearing on derivatives represents a particularly interesting case; the hearing, which was held before the Senate Committee on Agriculture, Nutrition, and Forestry, was called under the Committee's jurisdiction over commodities regulation and the Commodities Futures Trading Commission. Requesting testimony from the Secretary of the Treasury represented an unusual step, but highlighted the shifting jurisdictional boundaries and encroachment on issue territory that crisis events can cause. As Senator Saxby Chambliss noted:

> It is not often that the Secretary of the Treasury is called before the Ag Committee, but you have played an integral role thus far in dealing with this issue from a reform standpoint [...] It is imperative in my mind that the Senate Ag Committee should be engaged in the development of any legislation addressing financial regulation and, more specifically, derivatives. This Committee has a responsibility to ensure that the CFTC continues to effectively carry out its duties, including any new authorities and responsibilities Congress requires in the proposed financial regulatory reform legislation.[14]

These permeable jurisdictional boundaries are a prime example of the interplay

[10] Monetary Policy and the State of the Economy. Senate. 109th Congress, 2006. Full text available through the Government Publishing Office.

[11] Holding Banks Accountable: Are Treasury and Banks Doing Enough to Help Families Save Their Homes? Senate. 111th Congress, 2010. Full text available through the Government Publishing Office.

[12] The Federal Bailout of AIG. House. 111th Congress, 2010. Full text available through the Government Publishing Office.

[13] Over the Counter Derivatives Reform and Assessing Systemic Risk. Senate. 111th Congress, 2009. Full text available through the Government Publishing Office.

[14] Saxby Chambliss. Over the Counter Derivatives Reform and Assessing Systemic Risk. Senate. 111th Congress, 2009. Full text available through the Government Publishing Office.

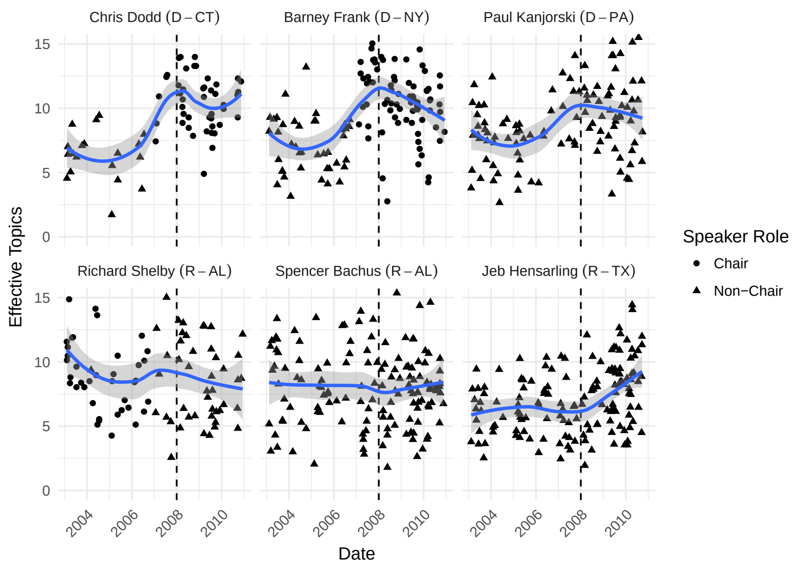
between cognitive constraints and strategic considerations highlighted throughout this paper. Though committees like the Senate Agricultural Committee ordinarily deal relatively little with financial policy and regulation, the importance of the 2007-2008 Financial Crisis and the role of derivatives in that event incentivized members like Chambliss to attempt to claim credit for particular regulatory reforms. As a result, Chambliss and others were likely more willing than usual to assume the accompanying cognitive costs involved in the information-gathering and policy-making process in an unfamiliar policy area.

We can conduct a similar analysis to the one described above on members of Congress. As Figure 4 highlights, members of Congress also vary substantially on the dimensionality of their conversation. Most notably, after party control of Congress changed hands in the 2006 elections, a number of members dramatically changed their discourse patterns. Chris Dodd and Barney Frank, the incoming chairmen of the Senate Banking Committee and the House Financial Services Committees, respectively, exhibited a particularly substantial increase in conversation dimensionality. Paul Kanjorski (the incoming chairman of the House Subcommittee on Capital Markets, Insurance, and Government-Sponsored Enterprises) displayed a similar shift. In all three cases, changes in conversation dimensionality appear to precede the onset of the financial crisis, suggesting that the changes in conversation patterns among these members are attributable to changes in leadership status rather than to the crisis.

As before, this basic pattern fits with our intuitions about the interplay between cognitive constraints, Congressional organization, and strategic incentives. Chris Dodd and Barney Frank, in particular, are members whose personal brands are strongly associated with regulatory reform and financial policy (e.g., through the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010). More generally, committee chairs like Dodd and Frank and senior members of the majority party also control the agenda and schedule for Congressional hearings, both through procedural mechanisms (opening/closing statements, control of parliamentary proceedings) and through informal control of the substantive hearing agenda. We should therefore expect members in these positions to be willing to engage more deeply with their preferred issues than other, less policy-focused representatives.

Interestingly, at least among leadership members pattern does not seem constant across parties. In general, Republicans and Democrats show no difference in conversation dimensionality ($p=0.71$, Welch's t-test). However, leadership Republicans like Richard Shelby (Chairman, Senate Banking Committee, 2003-2007) and Spencer Bachus (Ranking Member, House Financial Services Committee, 2007-2011) appear unaffected by either the 2007 change of partisan control of Congress or the onset of the Financial Crisis in 2008. Jeb Hensarling, who assumed chamber-level leadership duties after Republicans regained control of the House in 2011, does diversify his conversation slightly towards the end of the time series, but without additional data it is difficult to be sure if this trend is a small-sample artifact.

Figure 5: Effective topics values for selected members of Congress. Dashed lines represent January 1, 2008, the date at which the US economy officially entered a recession.



Since the dataset presented in this paper only covers one full change in partisan control, our ability to test the generality of this phenomenon is limited. However, one plausible explanation for the differences described above relates to the "partisan asymmetry" described in Grossmann and Hopkins [18]. According to the authors' argument, the major American parties are organized in distinct fashions: the Democratic Party, they argue, is "fundamentally a group coalition, [while] the Republican party can be most accurately characterized as the vehicle of an ideological movement."[18, 3] Upon assuming the chairmanships of their respective committees, Chris Dodd and Barney Frank both made statements supporting this general characterization of their party:

> As I have said previously, it is my intention to focus this Committee's attention on two fundamental objectives: first, strengthening our Nation's ability to keep our people and businesses as secure as possible against the risk of attack from those who wish us ill; and, second, expanding prosperity for businesses and consumers throughout our Nation.[15]
>
> I want to begin with an expression of disappointment, not in Chairman Bernanke, but in the business community and many of my conservative colleagues. I believe that we are at a very sensitive

[15] Christopher Dodd. Examining the State of Transit Security. Senate. 110th Congress, 2007. Full text available through the Government Publishing Office.

> point in the making of economic policy in this country [...] Many of us are prepared to work towards policies that are pro growth, that do take advantage of what you have when capital is allowed to reach its best level and find its greatest return, when technology can be fully taken advantage of, but only if we put in place public policies that make sure that is more fairly shared.[16]

In seeking to contrast themselves with their Republican predecessors, both Chris Dodd and Barney Frank chose to focus on income inequality as their primary concern. When given the opportunity to expand their discourse through committee leadership, then, members like Dodd and Frank were likely inclined to expand their discourse to include the various issues faced by members of disadvantaged groups. Republican leaders, by contrast, likely remained focused on traditional macroeconomic and business-related concerns, and were likely less inclined to use the procedural tools afforded committee leaders to expand their discourse. Testing this hypothesis more fully, however, would require a dataset covering a longer Congressional time series, and represents a direction for future research.

6. Conclusion

In summary, then, this paper offers two primary contributions. First, I outline and present a text-based approach to the study of cognitive patterns in political texts. As I argue throughout this paper, many of the most important open questions in political behavior and public policy involve an interplay between strategic behavior and cognitive limitations, which limit individuals' ability to pursue standard, rationally-generated strategies. In existing work, data limitations have prevented most researchers from examining these factors at an individual level, forcing major projects and research initiatives (e.g. the Comparative Agendas Project) to shift their focus to institutional-level patterns of behavior. Though this work has produced important insights, these limitations have prevented researchers from studying individual actors directly. Fortunately, new text-based data sources (e.g. transcript and speech data) offer new opportunities to examine individual-level distribution of attention to various policy areas, which we can exploit using newer modeling innovations and measurement strategies.

Second, I apply this measurement strategy to study patterns of attention distribution and issue engagement in American Congressional discourse on financial policy surrounding the 2007-2008 Financial Crisis. In particular, I find evidence for an expansion in conversation dimensionality following the crisis, as well as

[16] Barney Frank. Monetary Policy and the State of the Economy, Part 1. House. 110th Congress, 2007. Full text available through the Government Publishing Office.

systematic variation in issue engagement by speaker type and role. These kinds of findings expand our existing understanding of the interplay between cognitive limitations and strategic factors, giving us new insight into the early-stage information management and processing steps involved in the policymaking process.

[1] E Scott Adler and John D Wilkerson. *Congress and the politics of problem solving*. Cambridge University Press, 2013.

[2] Frank R Baumgartner and Bryan D Jones. *Agendas and instability in American politics*. University of Chicago Press, 2010.

[3] Frank R Baumgartner, Bryan D Jones, and Michael C MacLeod. The evolution of legislative jurisdictions. *The Journal of Politics*, 62(02):321–349, 2000.

[4] 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.

[5] Jonathan Bendor and Adam Meirowitz. Spatial models of delegation. *American Political Science Review*, 98(02):293–310, 2004.

[6] David M Blei. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84, 2012.

[7] David M Blei and Michael I Jordan. Variational methods for the dirichlet process. In *Proceedings of the twenty-first international conference on Machine learning*, page 12. ACM, 2004.

[8] David M Blei and John D Lafferty. Dynamic topic models. In *Proceedings of the 23rd international conference on Machine learning*, pages 113–120. ACM, 2006.

[9] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

[10] Amber E Boydstun, Shaun Bevan, and Herschel F Thomas. The importance of attention diversity and how to measure it. *Policy Studies Journal*, 42 (2):173–196, 2014.

[11] William G Chase and Herbert A Simon. Perception in chess. *Cognitive psychology*, 4(1):55–81, 1973.

[12] Matthew James Denny and Arthur Spirling. Assessing the consequences of text preprocessing decisions. Working paper, 2016.

[13] Sean Gailmard and John W Patty. Slackers and zealots: civil service, policy discretion, and bureaucratic expertise. *American Journal of Political Science*, 51(4):873–889, 2007.

[14] Sean Gailmard and John W Patty. *Learning while governing: expertise and accountability in the executive branch*. University of Chicago Press, Chicago, 2012.

[15] Thomas W Gilligan and Keith Krehbiel. Asymmetric information and legislative rules with a heterogeneous committee. *American Journal of Political Science*, pages 459–490, 1989.

[16] Justin Grimmer. A bayesian hierarchical topic model for political texts: Measuring expressed agendas in senate press releases. *Political Analysis*, 18(1):1–35, 2010.

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

[18] Matt Grossmann and David A Hopkins. *Asymmetric Politics: Ideological Republicans and Group Interest Democrats*. Oxford University Press, 2016.

[19] International Monetary Fund. World economic outlook - april 2009: Crisis and recovery. Technical report, Washington, DC, 2009.

[20] Bryan D Jones. Bounded rationality. *Annual review of political science*, 2 (1):297–321, 1999.

[21] Bryan D Jones and Frank R Baumgartner. *The politics of attention: How government prioritizes problems*. University of Chicago Press, Chicago, IL, 2005.

[22] Bryan D Jones and Frank R Baumgartner. From there to here: Punctuated equilibrium to the general punctuation thesis to a theory of government information processing. *Policy Studies Journal*, 40(1):1–20, 2012.

[23] David C King. *Turf wars: How Congressional committees claim jurisdiction*. University of Chicago Press, 1997.

[24] Wai Fung Lam and Kwan Nok Chan. Policy advocacy in transitioning regimes: Comparative lessons from the case of harbour protection in hong kong. *Journal of Comparative Policy Analysis: Research and Practice*, pages 1–18, 2015.

[25] Arthur Lupia and Mathew D McCubbins. Learning from oversight: Fire alarms and police patrols reconstructed. *Journal of Law, Economics, & Organization*, pages 96–125, 1994.

[26] Thomas E Mann and Norman J Ornstein. *The broken branch: How Congress is failing America and how to get it back on track*. Oxford University Press, 2006.

[27] James G March. *Primer on decision making: How decisions happen*. Free Press, New York, NY, 1994.

[28] Andrew Kachites McCallum. Mallet: A machine learning for language toolkit. http://mallet.cs.umass.edu, 2002.

[29] Mathew D McCubbins. The legislative design of regulatory structure. *American Journal of Political Science*, pages 721–748, 1985.

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

[31] Mathew D McCubbins, Roger G Noll, and Barry R Weingast. Administrative procedures as instruments of political control. *Journal of Law, Economics, & Organization*, 3(2):243–277, 1987.

[32] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

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

[34] 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.

[35] Fred Paas, Alexander Renkl, and John Sweller. Cognitive load theory and instructional design: Recent developments. *Educational psychologist*, 38 (1):1–4, 2003.

[36] Radim Řehůřek and Petr Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May 2010. ELRA. `http://is.muni.cz/publication/884893/en`.

[37] Margaret E Roberts, Brandon M Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G Rand. Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4):1064–1082, 2014.

[38] Herbert A Simon. Human nature in politics: The dialogue of psychology with political science. *American Political Science Review*, 79(02):293–304, 1985.

[39] Herbert A Simon. *The sciences of the artificial.* MIT Press, Cambridge, MA, 1996.

[40] Charles III Stewart and Jonathan Woon. Congressional committee assignments, 103rd to 112th congresses. 2011.

[41] John Sweller. Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, 4(4):295–312, 1994.

[42] John Sweller. Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational psychology review*, 22(2):123–138, 2010.

[43] Jeffery C Talbert, Bryan D Jones, and Frank R Baumgartner. Nonlegislative hearings and policy change in congress. *American Journal of Political Science*, pages 383–405, 1995.

[44] Yee Whye Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Hierarchical dirichlet processes. *Journal of the american statistical association*, 2012.

[45] Amos Tversky and Daniel Kahneman. The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458, 1981.

[46] Hanna M Wallach, David M Mimno, and Andrew McCallum. Rethinking lda: Why priors matter. In *Advances in neural information processing systems*, pages 1973–1981, 2009.

[47] Samuel Workman, Bryan D Jones, and Ashley E Jochim. Information processing and policy dynamics. *Policy Studies Journal*, 37(1):75–92, 2009.

Supporting Information

Robert Shaffer[1]

University of Texas, Austin

AppendixA. Effective Topics Derivation and Discussion

Observed information entropy of a categorical random variable X is defined as follows:

$$H(X) = \frac{1}{log(n)} \sum_{i=1}^{n} p_i log\Big(\frac{1}{p_i}\Big)$$

With n the number of bins, p_i the observed cell proportions, and $0 < p_i < 1$ for all p_i.

As mentioned in-text, informational entropy is highly non-linear, and for interpretive purposes we may wish to place entropy on some linear scale. One such scale is the "effective topics" scale, or the *number of equally proportioned-bins* (for fixed total number of bins n) that would have produced the same informational entropy as the original dataset. Transforming entropy to effective topics places entropy on a linear scale with respect to an intuitive quantity, allowing readers to interpret the statistic more easily.

We can formally derive the effective topics transformation as follows. Suppose we make a set of observations on a categorical random variable X with n bins. Further suppose $H(X) = \eta$. Given this sample information, our goal is to find

Email address: `rbshaffer@utexas.edu` (Robert Shaffer)

a theoretical alternative set of cell proportions Y such that:

$$H(Y) = \eta \tag{A.1}$$

$$\sum_{i=1}^{k} m_i = 1 - \epsilon \tag{A.2}$$

$$m_x = m_y \forall x, y \in \{1, 2...k\} \tag{A.3}$$

$$\sum_{i=k+1}^{n} m_i = \epsilon \tag{A.4}$$

$$m_a = m_b \forall a, b \in \{k+1, ...n\} \tag{A.5}$$

With m_i the observed cell proportions, k an unknown positive integer, and $0 < \epsilon < 1$. Conditions (A.2-A.5) imply that $m_i = \frac{1-\epsilon}{k} \forall \{1, 2, ...k\}$, and $m_j = \frac{\epsilon}{n-k} \forall \{k+1...n\}$. Note that bin ordering can be rearranged without loss of generality.

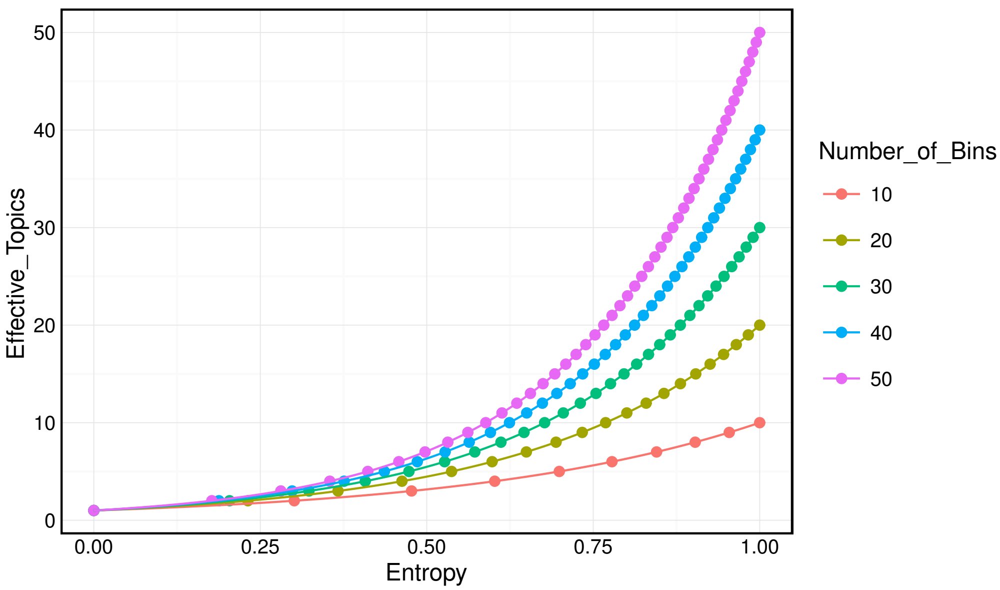
Taking the limit as ϵ goes to 0 from the positive side (the case in which $n-k$ residual cell probabilities are arbitrarily close to 0), we can rewrite $H(Y)$ as follows:

$$\begin{aligned}
\lim_{\epsilon \to 0^+} H(Y) &= \lim_{\epsilon \to 0^+} \frac{1}{log(n)} \Big[\sum_{i=1}^{n} m_i log\Big(\frac{1}{m_i}\Big)\Big] \\
&= \lim_{\epsilon \to 0^+} \frac{1}{log(n)} \Big[\sum_{i=1}^{k} m_i log\Big(\frac{1}{m_i}\Big) + \sum_{j=k+1}^{n} m_j log\Big(\frac{1}{m_j}\Big)\Big] \\
&= \lim_{\epsilon \to 0^+} \frac{1}{log(n)} \Big[\frac{k}{k}(1-\epsilon) log\Big(\frac{k}{1-\epsilon}\Big) + \frac{n-k}{n-k} \epsilon log\Big(\frac{n-k}{\epsilon}\Big)\Big] \\
&\to \frac{log(k)}{log(n)}
\end{aligned}$$

Solving this expression for k gives $k = n^{\eta}$, giving the result from the body of the paper.

Importantly, note that this result is only valid for values of η that return an integer value of k (for all other values of η, there is no solution to the problem posed in this Appendix). From a more informal standpoint, however, we can view the function η_n as an interpolation between integer values of n (see Figure A.6 below for examples). As a result, the same basic intuition that underlies this proof can be extended to values of η that return non-integer values of k.

Figure A.6: Simulated effective topics and varying numbers of bins.



AppendixB. Hearing Parser and Metadata Association Algorithm

As mentioned in-text, the GPO delivers hearing transcripts as plain-text files, with no embedded metadata. Moreover, the formatting and ordering of information is not consistent across the transcripts in the GPO's dataset. As a result, parsing these transcripts and linking the parsed files to individual-level metadata represents a substantial task in itself.

For the purposes of this project, we developed a specialized regular expression-based parser, which relied on heuristic observations regarding the GPO's formatting standards, committee membership information drawn from Stewart's dataset, and whatever hearing-level metadata were available on the GPO's website. For details regarding the parser protocol, see the project replication code; however, the algorithm can roughly be summarized as follows:

Require: Hearing transcripts X.
Require: Stewart's Congressional committee membership data C.
Require: GPO hearing witness data W.

for $x \in X$ **do**
    Extract hearing-level metadata M from x
    Segment x into sessions J.
    Strip all non-spoken materials from x.

    for $j \in J$ **do**
        Segment j into statements K.

        for $k \in K$ **do**
            Extract the last name a of each speaker from k.
            if $a \in C_x$ **then**
                Assign C_a to k.
            else if $a \in W_x$ **then**
                Assign W_a to k.
            else if $a \in M \& a \in C$ **then**
                Assign C_a to k.
            elseAssign NA to k.
            end if

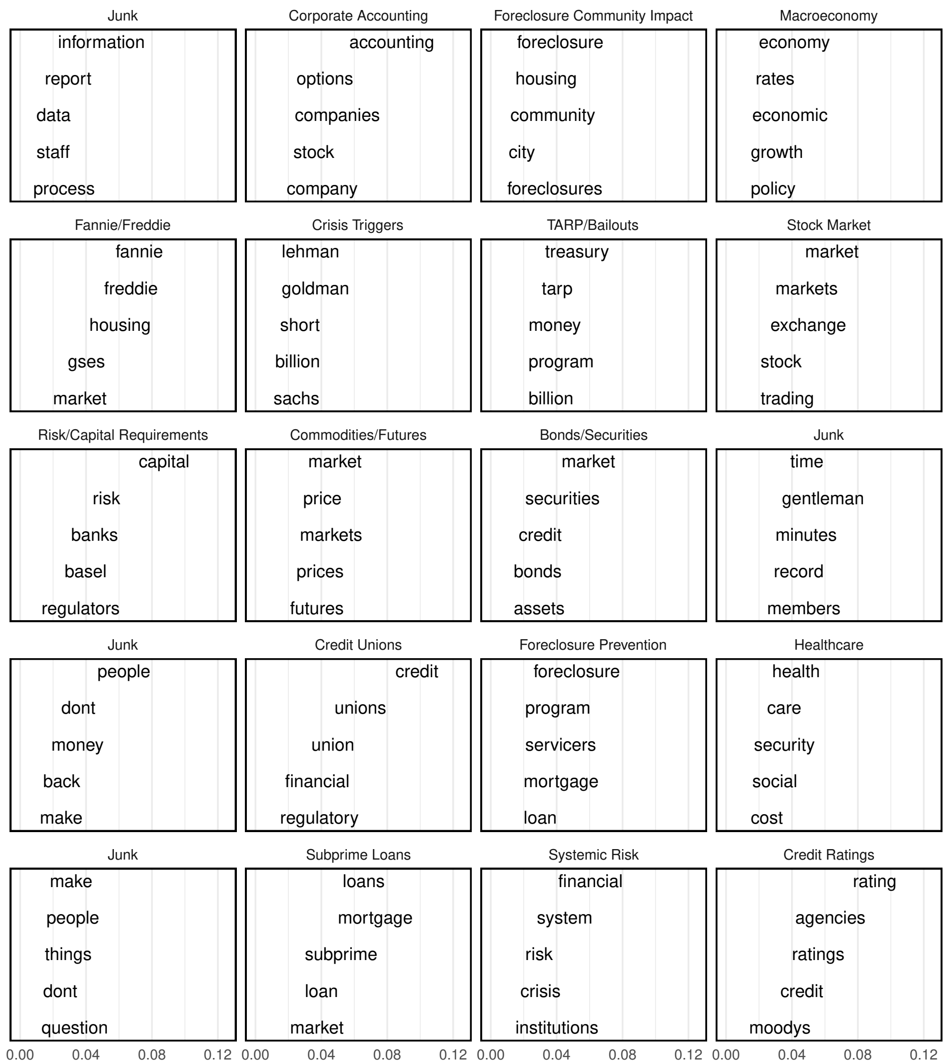
        end for
    end for
end for

AppendixC. Full LDA Topics List

The five highest probability words for each topic in the 40-topic model used in-text are given in Figures C.7 and C.8. As noted in-text, some topics cover non-substantive usage areas such as parliamentary procedure and presentation of evidence. For simplicity, non-substantive topics are titled "junk" topics.

Figure C.7: Top probability words for 40-topic model.

Figure C.8: Top probability words for 40-topic model.



AppendixD. Robustness

As mentioned in-text, unsupervised text analysis applications involve an array of pre-processing and parameter selection steps, many of which are difficult to defend ex ante. Testing robustness of conclusions to these choices is therefore an important analytical step.

In this Appendix, I present two sets of robustness results. First, I examine the robustness of in-text conclusions to the choice of K (the LDA dimensionality parameter). Second, I focus on the 40-topic model used in-text, and and examine robustness of in-text conclusions to the inclusion or exclusion of "junk" topics (i.e. those topics do not appear to be related to substantive policy areas).

AppendixD.1. Varying K

Replications at $K \in \{20, 25, ...100\}$ for Figures 2 and 3 are given in Figures D.9 and D.10. At all values tested, the basic conclusions given in-text remain consistent. All subgroups display a noticeable spike in conversation dimensionality following the onset of the crisis, followed by a slow decline. Members of the Senate display higher effective topics values than members of the House, and leadership members display consistently higher effective topics values than members of other subgroups. Varying K does induce an intercept shift in the underlying effective topics data, suggesting that the absolute scale given in text is essentially arbitrary. However, within-time series and cross-subgroup relative differences are consistent across values of K.

Figure D.9: Smoothing spline fit to effective topics values calculated on statements aggregated to the speaker-hearing level and divided by chamber. Dashed line indicates January 1, 2008, the date at which the US economy officially entered a recession.

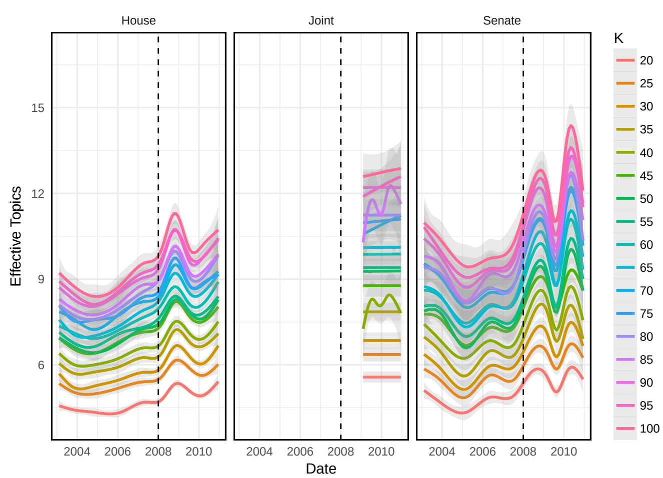
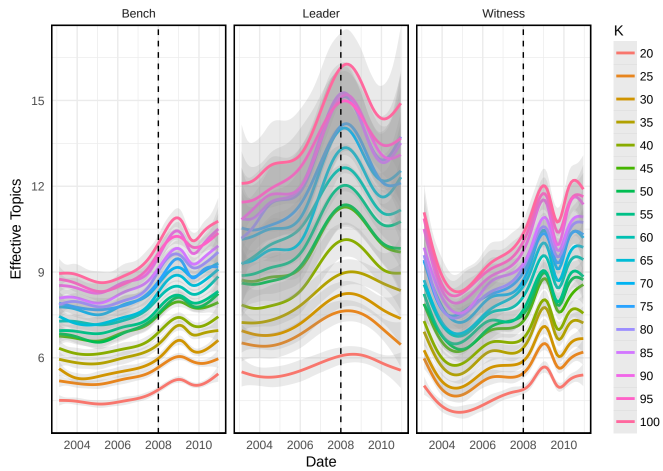


Figure D.10: Smoothing spline fit to effective topics values, aggregated to the speaker-hearing level and divided by speaker type. Dashed line indicates January 1, 2008, the date at which the US economy officially entered a recession.



AppendixD.2. Removing "Junk" Topics

Replications with "Junk" topics removed for Figures 2 and 3 for the 40-topic model used in-text are given in Figures D.11 and D.12. Topics were identified as "junk" (i.e. non-substantive) through a double-coding and reconciliation process. To remove "junk" topics, "junk" bins were removed from each proportion vector, and the remaining proportions were re-normalized. This process left a total of 31 non-junk topics in each speaker-hearing observation.

As shown below, removing "junk" topics did not affect the substantive results given in-text. In this specification, the post-crisis spike in conversation dimensionality remains constant, with the largest effect observed among members of the Senate. The only group substantially affected by the removal of "junk" topics are leadership members (though leadership members still address a larger range of topics than witnesses and their backbench counterparts at most time periods in the dataset. This result seems intuitively plausible. Since leadership members discuss procedural matters more frequently than other speakers, a larger proportion of their verbiage should be devoted to non-policy discussions. As a result, their discourse is likely to be disproportionately affected by the removal of non-policy language.

Figure D.11: Smoothing spline fit to effective topics values, aggregated to the speaker-hearing level and divided by chamber. Dashed line indicates January 1, 2008, the date at which the US economy officially entered a recession.

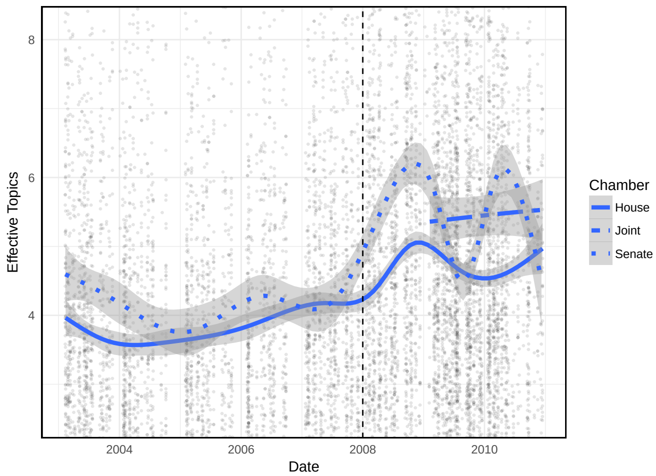


Figure D.12: Smoothing spline fit to effective topics values, aggregated to the speaker-hearing level and divided by speaker type. Dashed line indicates January 1, 2008, the date at which the US economy officially entered a recession.