Cognitive Load and Issue Engagement in Congressional Discourse

Robert Shaffer[1]

University of Texas, Austin

Abstract

Like all human actors, politicians possess limited cognitive capacity. In ordinary interactions, this limitation discourages political decision-makers from addressing high-dimensional policy problems unless incentivized to do so by exogenous "focusing events." Public policy researchers have documented this pattern extensively, and have argued that cognitive constraints help explain the "stick-slip" dynamics that characterize macro-level policymaking. However, data and measurement limitations have prevented these studies from examining individual-level information processing patterns.

In this paper, I develop a text-based approach designed to measure diversity of attention at an individual level, which I apply to an original dataset of Congressional hearing transcripts surrounding the 2008-2009 Financial Crisis. I find that individual speakers engaged with a more diverse set of topics during the crisis than before its onset, and became more focused as the crisis subsided.

1. Introduction

In decision-making scenarios, a key challenge human actors face is the problem of managing *issue dimensionality.* When deciding on consequential matters, actors grapple with a dizzying array of information. 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 effects, including inflation, unemployment, business debt pricing, and homeownership.

Usually, cognitive limitations prevent individuals from considering all aspects of a particular issue [14, 27]. Political institutions follow this same pattern. Like the individuals that compose them, legislative bodies like the American

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

Congress can only focus on a few ideas at once, leaving policy in most areas to languish until a crisis point is reached. Researchers have argued that these cognitive constraints explain the disproportionate allocation of attention that characterizes high-level policymaking, in which attention to individual issues languishes for long periods that are "punctuated" by brief spikes of increased interest and engagement [e.g. 15, 1, 16].

The attention allocation pattern described above is well-documented at the aggregate level, but few existing studies have examined information processing patterns among individual decision-makers. This disconnect is troubling; since many of the existing explanations for aggregate-level policymaking patterns are based on individual-level cognitive phenomena, providing evidence for the presence of these effects is a critical analytic step. Moreover, the lack of focus on individual-level patterns leaves important questions regarding individual-level allocation of attention unanswered. In particular, how does expertise or familiarity with a problem area affect an individual's willingness to raise a broad set of problem aspects? And, are particular types of actors more willing to raise a broad set of issue dimensions than others?

In this paper, I propose a text-based measurement approach designed to address these questions. In recent decades, institutions like the US Congress have made individual-level text data (e.g. hearing transcripts) increasingly available. As I argue, topic models and other unsupervised dimensionality reduction tools are well-suited for detecting changes in allocation of attention. I outline an approach based on these methods, and apply this approach to examine patterns in conversation dimensionality in the US Congress from 2004 to 2011. Overall, I find that dimensionality 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 dataset, with experts and leadership members engaging more deeply with relevant policy problems than other speakers.

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 [27, 28] argues, individual behavior in decision-making settings can be best described as intendedly rational. Though human actors usually attempt to pursue goal-directed, utility-maximizing patterns of behavior, their ability to follow these strategies is constrained. In particular, individuals possess limited ability to consider and compare the relevant dimensions of various problems, creating an "oversupply of information" [34] 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. 32, 18, 14].

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" [24]. 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 [30]. 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" [8] or eliminate non-germane problem aspects [31] in order to ease individual-level cognitive demands.

Translated to the political domain, these cognitive phenomena produce an intuitive set of behavioral predictions. Like other human actors, politicians tend to avoid addressing issues that involve a heavy cognitive load. For a concrete example, consider legislative oversight. As McCubbins and Schwartz [21] famously argue, oversight activity in the US Congress 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 broad set of issue areas. 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 [21, 175].

While the "rationality" of crisis-based issue management is debatable, fire alarm-type oversight is clearly cognitively appealing. By delegating oversight authority and establishing broad "framework"-style legislation, lawmakers can focus their energy on a narrow set of important or familiar problem areas, and avoid the heavy cognitive load associated with a broad oversight agenda [see also, e.g. 17, 20, 22]. When "police patrol"-style oversight is unavoidable, politicians tend to favor routinized, automatic processes which are slow to adjust to changes in external conditions [1, 15, 16]. As a result, policymaking more generally tends to follow a "stick-slip" pattern, in which legislators and bureaucrats allow policy in particular areas to languish until a crisis point is reached [2, 1].

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 test hypotheses related to individual-level behavior (as opposed to aggregate-level patterns). Largely, this limitation results from data constraints. Since major datasets like the Comparative Agendas Project (CAP) code distribution of attention at the hearing or bill level, hypotheses regarding individual-level characteristics cannot be tested using these data. This disconnect is troubling; intuitively, we should expect some individuals (e.g. those with a leadership role or preexisting expertise in a particular policy area) to be more willing to engage with a broad set of issues than their less-incentivized counterparts. Datasets like CAP that focus on aggregate-level agenda-setting patterns ignore these patterns, and leave us unable to test hypotheses regarding individual-level characteristics.

For the remainder of this paper, I focus on three such hypotheses, which I outline below:

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 2008-2009 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. For broad-ranging events like the Financial Crisis, we should expect members of virtually all subgroups to engage with a more diverse set of of ideas; however, for smaller-scale or more issue-specific events, we might expect members of some subgroups to be affected more noticeably than others.

2.2.2. Institutional Investment

Based on existing work, we should expect most individuals to adopt information-processing strategies that impose a low cognitive load. However, we should also expect individual-level characteristics to affect a given person's willingness to shoulder a heavy cognitive burden. In the context of Congressional discourse, members with broader constituencies and a more national profile are likely to be more invested in the success of a given policy program, and more willing to engage deeply with the policy problems that program addresses. Committee chairmen and other leadership members, for example, have a personal stake in the committee's success, and and are likely more willing to press a more expansive (and more cognitively taxing) view of a given committee's agenda. Similarly, compared with members of the House, Senators possess broader constituencies and face greater incentives to explore policy problems more deeply.

In many cases, the organization of Congressional hearings reinforces this expectation. Leadership members frequently make broad opening and closing statements before the start of each hearing, which set the agenda and the general themes for the day's discussion. These statements provide additional opportunities for leadership members to speak in a wide-ranging fashion, and allow them to discuss a broader range of ideas than other hearing participants.

2.2.3. 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).

Again, 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.

2.3. Discussion

Before proceeding, two caveats are in order. First, some of the factors described above should affect both individual- and aggregate-level information processing patterns. For example, large-scale "focusing events" like the 2008-2009 Financial Crisis should incentivize virtually all individuals to engage with a broader set of issues, inducing effects that are observable at any level of analysis. However, in other cases, the effects we expect to see are contingent on individual-level characteristics. Intuitively, we should expect traits that ease the cognitive load imposed by complex tasks (or incentivize individuals to shoulder that load) to produce a more diverse policy discourse. However, since these traits are individual-level characteristics, we cannot test these expectations with aggregate-level data.

Second, not all variation in conversation dimensionality is driven by cognitive effects. Institutional rules 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 exogenous events like the Financial Crisis should be directly attributable to cognitive responses.

3. Dataset Creation and Measurement

3.1. The Financial Hearings Corpus

To test these hypotheses, I examine patterns of individual-level issue engagement in an original dataset of Congressional hearing transcripts surrounding the 2008-2009 Financial Crisis. 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. These hearings were selected based on their relationship to known policy areas affected by the Financial Crisis [see, e.g., 13], and represent a substantively important subset of Congressional discourse.

Since the hypotheses posed in this paper largely deal with individual-level characteristics, I then converted each hearing transcript into a series of individual-level statements. Using a series of custom Python tools, I segmented each transcript by speaker and linked each statement with metadata (e.g. party, speaker type, speaker seniority) drawn from the GPO's website and from Stewart and Woon [29]'s committee membership data.[1] For most committees, the GPO's coverage begins in 2004; however, at time of publication Stewart and Woon [29]'s committee membership data was not available beyond the 112th Congress, restricting the time period for this dataset to 2004-2011. In total, this dataset consists of 582 hearings drawn from 23 distinct committees in both chambers of Congress.

3.2. Issue Dimensionality in Text Documents

3.2.1. Existing Approaches

In an ideal world, a researcher interested in measuring issue engagement in this corpus (or a similar dataset) would follow a two-step procedure. First, the researcher would generate a conceptual inventory of all possible topics that might be raised by the speakers under study. Second, the researcher would code each speaker's statements according to the proportion of time spent on each topic, producing a compositional vector $p = [e_1, e_2, \ldots\ e_n]$ with n topics, $e_i \geq 0 \forall i \in \{1, 2, \ldots n\}$, and $\sum_{i=1}^{n} e_i = 1$. As shown in the following section, compositional vectors of this kind can be easily used used to measure diversity of conversation across various speakers, allowing the researcher to test hypotheses like those I describe in Section 2.

Historically, major datasets like the Comparative Agendas Project have relied on hand-coding procedures to convert documents to compositional vectors of

[1] 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.

this sort. Unfortunately, coding committee transcripts and other large data sources by hand is impractical. Parsing large documents with many speakers and many topics of interest is too labor-intensive for human readers, forcing researchers to restrict themselves to higher-level organizational priorities. For example, 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.

Machine-assisted methods, by contrast, are more promising. Over the last several decades, computer scientists and statisticians have developed an array of methods designed to extract latent themes or ideas from textual corpora. Prominent approaches in this area include latent Dirichlet allocation (LDA) and its many variants and extensions [e.g. 6, 4, 5, 10, 26], and more recent deep-learning based approaches such as Mikolov et al. [23]'s word2vec. Estimation and modeling details vary across these modeling approaches, but all essentially attempt to reduce high-dimensional text 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.

3.2.2. A Text-as-Data Alternative

To convert the hearings data described above to a machine-interpretable dataset, I began by transforming 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, 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

Table 1: Pre-processing specification.

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

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

words that occur in a few documents sacrifices little analytical leverage while easing computational burden.[2] 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. Using this cleaned dataset, I then fit a 40-topic latent Dirichlet allocation (LDA).[3] In its most basic form, LDA is a three-level Dirichlet-Multinomial hierarchical model, which treats documents as mixtures of latent "topics" (probability 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 [3]:

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

[2] In the broader literature, dropping larger sets of terms (e.g. terms that occur in fewer than 1% of all documents) is common [11]. As Denny and Spirling [9] 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 previously, 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.

[3] With the model fit via a Gensim wrapper for MALLET [25, 19] and the asymmetric prior setup described in Wallach et al. [33]. For a more detailed description of the LDA estimation and parameterization steps, see e.g., Blei [3].

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.[4]

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 later in this paper.[8]

3.2.3. Measuring Dimensionality

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), and normalized each summed vector ($n \approx 10,000$). To measure breadth of issue engagement in these speaker-hearing topic proportion vectors, I adopted and extended the approach suggested in Boydstun et al. [7], which suggests *informational entropy* as an appropriate measure of attention diversity. Informational entropy is defined as follows:

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

Where P is a topic proportion vector with n topics, and each item $p_i \forall i \in \{1, 2, ...n\}$ denotes the proportion of a given speaker's verbiage devoted to the i^{th} topic.

From a mathematical standpoint, informational entropy can be viewed 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 (i.e. $p_j = 1$ and $p_i = 0 \forall i \neq j$ for some $j \in \{1, 2, ...n\}$), while an informational entropy of $\eta = 1$ indicates that the vector P splits its verbiage evenly

[4] See Appendix D for details.

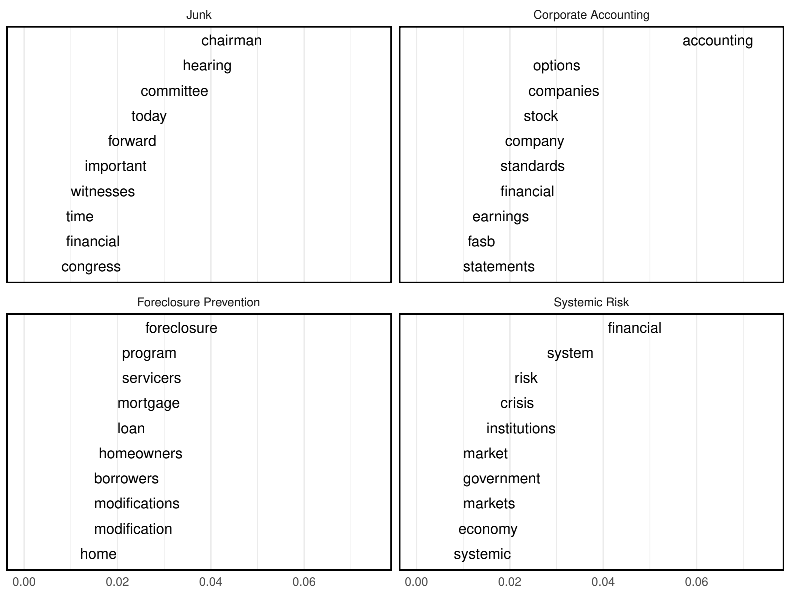
[5] Specifically, two coders independently read the 20 highest-probability words from each topic and the ten documents (statements) from each topic that contained the highest proportion of their words drawn from each topic. Then, each coder assigned labels to each topic. Finally, coders reconciled their 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. Terms are positioned such that the left edge of the term indicates the probability of drawing that term for a given topic.



across all topic values (i.e. $p_i = \frac{1}{n} \forall i$). This statistic therefore provides a natural way to measure the extent to which a particular speaker "focuses" his or her conversation on a particular idea.

Unfortunately, as Boydstun et al. [7] show, the informational entropy statistic is highly non-linear, making comparisons between various entropy values difficult. To address this issue, I 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.[9] 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,

[9] See Appendix A for proof and discussion

and allows readers to compare differences in a linearized fashion.

With this procedure, comparability of topics is a potential concern. Statistics like entropy (and the derived effective-topics transformation) implicitly assume that 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.[10]

3.3. Summary

Overall, then, the measurement strategy used 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 important modeling choices and validation steps, which I describe in detail in-text. However, the basic strategy is fairly straightforward. Modern model-fitting and data cleanup tools make estimation and model validation reasonably straightforward. 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.

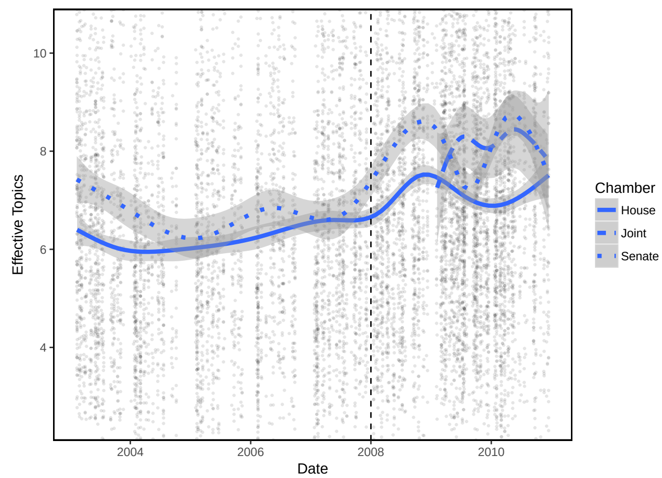
4. Issue Dimensionality in Congressional Discourse

4.1. Aggregate-Level Issue Dimensionality

As described in Hypothesis 1, the first and strongest relationship we should expect to observe in this dataset is a sharp expansion in conversation dimensionality following the onset of the Financial Crisis. Defining the start of the crisis

[10] See Appendix D for details.

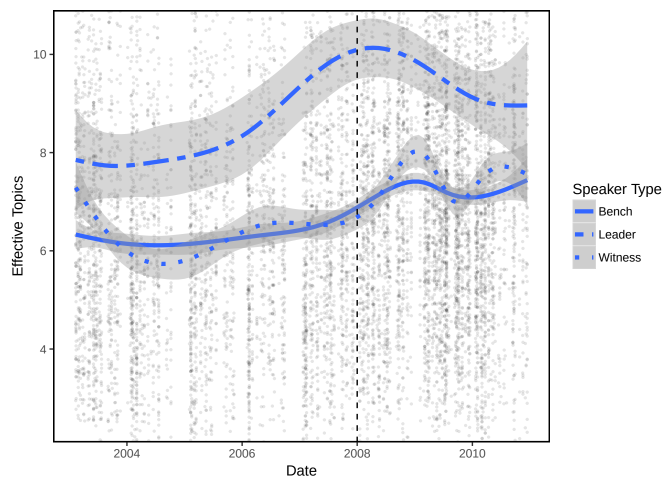
Figure 2: Effective topics values for combined speaker-hearing proportion vectors, separated by the chamber in which the hearing was held. Dashed line indicates January 1, 2008 (the official start of the Great Recession).



as January 1, 2008 (the date at which the US economy officially entered a recession), conversation dimensionality in financial 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. Unsurprisingly, all subgroups experience a similar post-crisis spike and subsequent decline in conversation dimensionality, suggesting that the basic cognitive impact of the crisis event was roughly comparable across all segments of the dataset. As predicted in Hypothesis 2, leadership members exhibit consistently higher effective topics values their backbench counterparts (9.0 effective topics versus 6.8; $p < 10^{-16}$, Welch's t-test). Contrary to Hypothesis 3, witnesses only discuss a slightly more diverse set of topics than backbench members of Congress (6.9 effective topics versus 6.8; $p \approx 0.01$, Welch's t-test). However, as shown in the following section, we can still detect patterns consistent with Hypothesis 3 by conducting within-group comparisons, suggesting that the positive relationship between expertise and breadth issue engagement predicted by Hypothesis 3 is restricted to witnesses with particularly broad issue expertise.

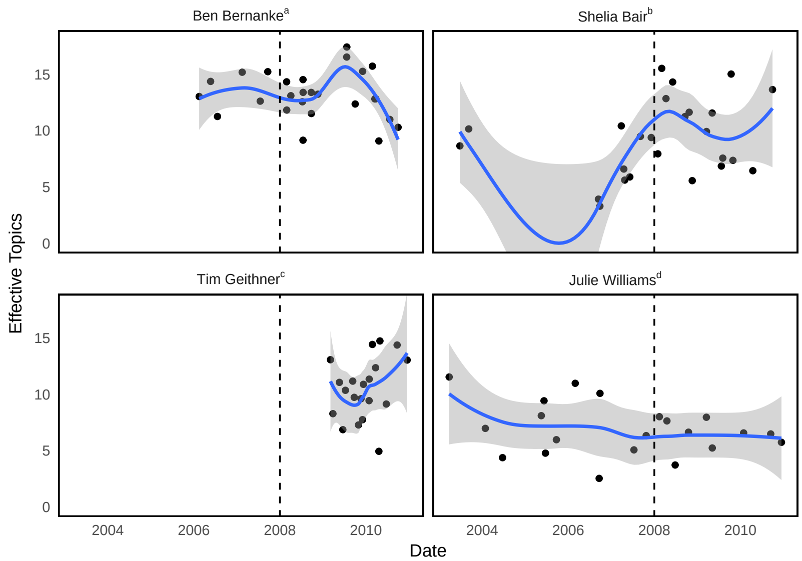
Figure 3: Effective topics values for combined speaker-hearing proportion vectors, separated by speaker type. "Leadership" members are defined as those members holding a committee chairmanship. Dashed line indicates January 1, 2008 (the official start of the Great Recession).



This basic pattern seems reasonable, and fits with the theoretical predictions given at the outset of this paper. Like many high-level policy problems, financial policy is complex and multidimensional, encouraging speakers to turn to heuristic devices (e.g. party platforms and expert recommendations) to avoid assuming a heavy cognitive burden. During crisis periods, familiar heuristic devices break down, forcing individuals to explore problem aspects and approaches outside of preexisting paradigms [14]. However, the incentive to dive more deeply into a particular problem area is likely not constant across all speakers. Leadership members and Senators, for example, usually possess larger and more diverse constituencies than their backbench and House counterparts, and possess a greater incentive to address diffuse policy problems such as macroeconomic events. Leadership members, in particular, also possess institutionally-defined opportunities to speak on a wide range of topics (e.g. opening and closing statements), giving them additional opportunities to speak broadly.

The pattern of decline following the end of the crisis also fits with this story. 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 fully 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.

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)

4.2. Individual-Level Attention Patterns

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

For starters, 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 the basic pattern described in Hypothesis 3. 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.[11] 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[12], the bailout of AIG[13] - both relatively focused topics - as well as broader ideas such as derivatives and their impact on systemic risk[14]. 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 that crisis events can provoke. 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.[15]

---

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

[12] 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.

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

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

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

Consistent with Hypotheses 1 and 2, the importance of the 2008-2009 Financial Crisis and the role of derivatives in that event incentivized institutionally-invested members like Chambliss to claim credit for regulatory reforms outside their natural issue jurisdiction. At least during this time period, then Chambliss and his colleagues were likely more willing than usual to assume the cognitive costs involved in the information-gathering and policy-making process in an unfamiliar policy area.

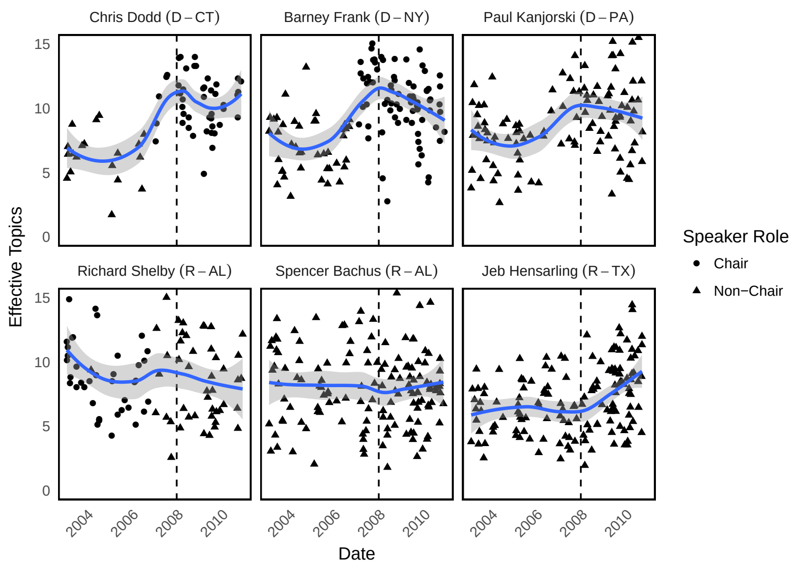
We can conduct a similar analysis to the one described above on other members of Congress. As Figure 4 highlights, members of Congress also vary substantially on the dimensionality of their conversation. Most notably, after party control of Congress changed hands in the 2006 elections, a number of members 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. Consistent with Hypothesis 2, changes in conversation dimensionality appear to precede the onset of the 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, the institutional investment effects predicted by Hypothesis 2 do 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.

Since the dataset presented in this paper only covers one full change in par-

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.



tisan 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 [12]. 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."[12, 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.[16]

> 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 point in the making of economic policy in this country [...] Many of

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

> 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.[17]

If we accept Grossmann and Hopkins [12]'s argument, these partisan differences fit naturally with the institutional investment effects predicted in Hypothesis 2. When given the opportunity to expand their discourse through committee leadership, Democratic members like Dodd and Frank appeared willing to expand their discourse to include the income inequality and other issues faced by 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.

5. Conclusion

In summary, this paper offers two primary contributions. First, I outline a text-based approach to the study of issue engagement in political discourse. As I argue throughout this paper, many of the most important theoretical developments in the public policy literature rely on individual-level cognitive features to explain observed phenomena. 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 directly testing hypotheses related to individual-level factors. Fortunately, new text-based data sources (e.g. hearing transcripts) offer new opportunities to examine individual-level distribution of attention to various policy areas, allowing researchers to test these kinds of hypotheses more directly.

Second, I apply this approach to study patterns of attention distribution and issue engagement in American Congressional discourse. Using an original dataset of Congressional hearings on financial policy, I find evidence for an expansion in conversation dimensionality surrounding the 2008-2009 Financial Crisis. I also find systematic variation in diversity of conversation by speaker type and role, with institutionally-involved speakers (e.g. leadership members and Senators) discussing a noticeably larger set of counterparts than their less invested

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

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

[2] 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.

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

[4] 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.

[5] 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.

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

[7] 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.

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

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

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

[11] 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.

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

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

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

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

[16] 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.

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

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

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

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

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

[22] 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.

[23] 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.

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

[25] 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`.

[26] 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.

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

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

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

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

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

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

[33] 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.

[34] 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.

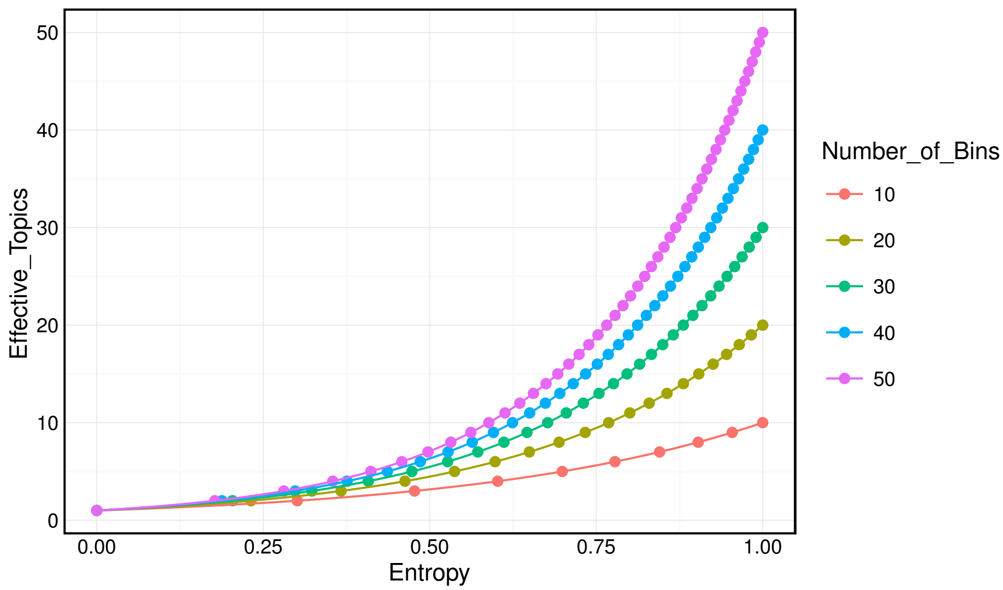
Taking the limit as ϵ goes to 0 from the positive side (the case in which $n-k$ residual cell probabilities are arbitrarily close to 0), we can rewrite $H(Y)$ as follows:

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

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, I 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

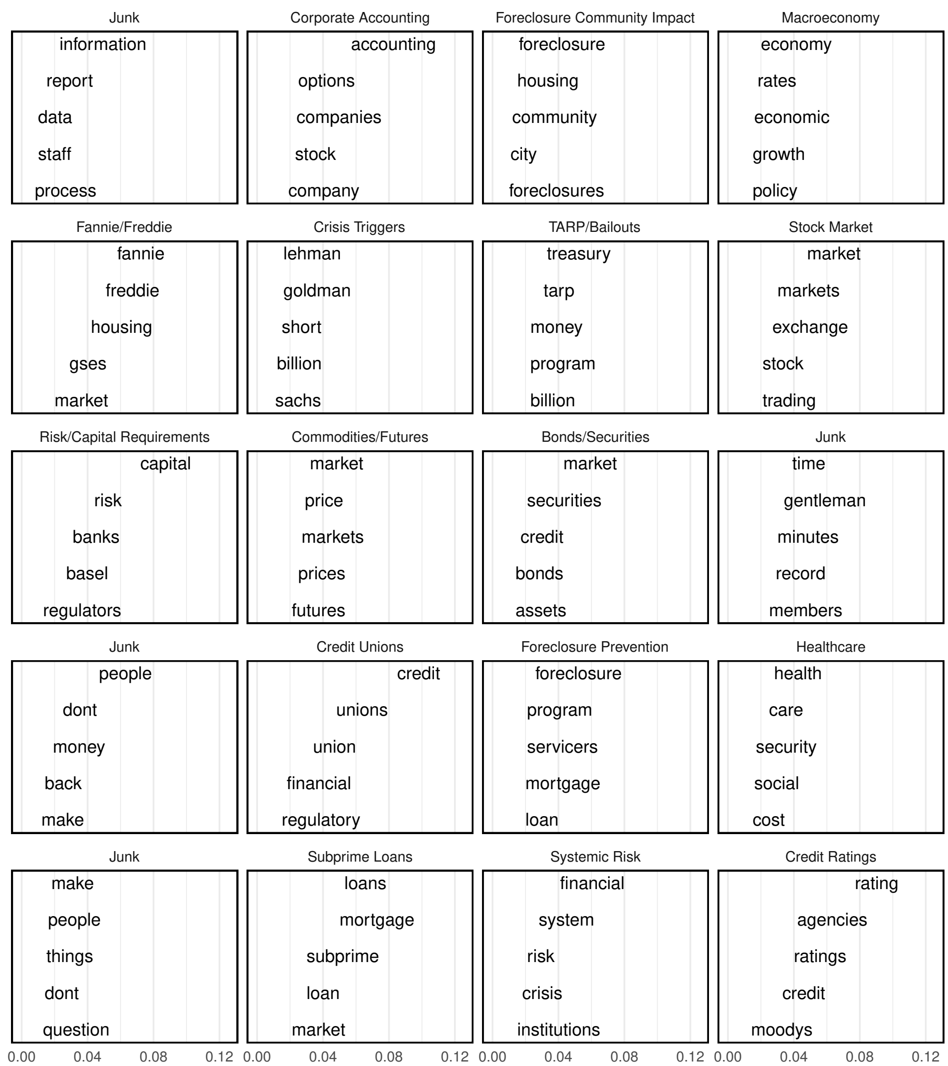
        end for
    end for
end for

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.

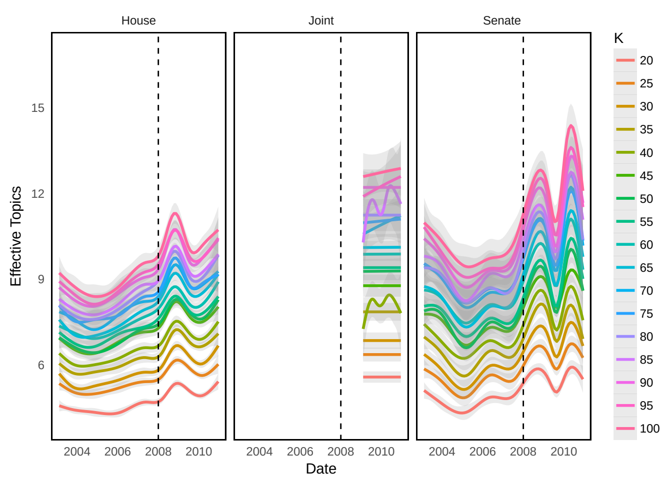
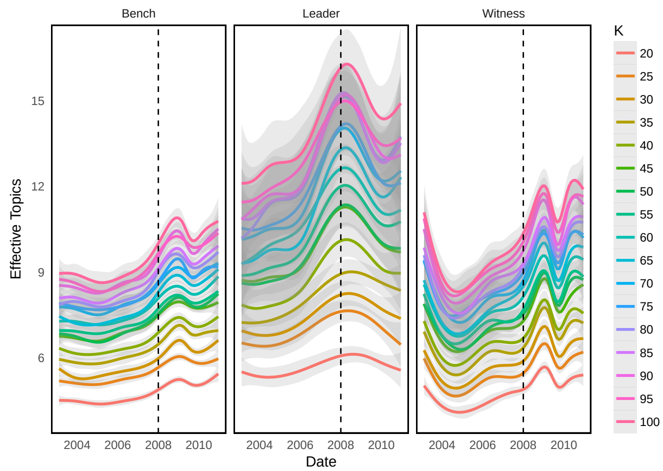


Figure D.10: Smoothing spline fit to effective topics values, aggregated to the speaker-hearing level and divided by speaker type. Dashed line indicates January 1, 2008, the date at which the US economy officially entered a recession.



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.

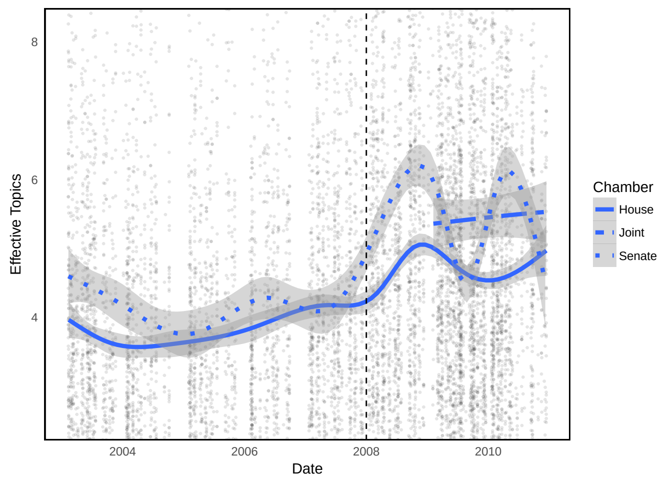


Figure D.12: Smoothing spline fit to effective topics values, aggregated to the speaker-hearing level and divided by speaker type. Dashed line indicates January 1, 2008, the date at which the US economy officially entered a recession.