Measuring Representational Style in the House: The Tea Party, Obama and Legislators' Changing Expressed Priorities

Justin Grimmer

October 30, 2014

Abstract

Legislators use public communication to define the type of representation they provide constituents. This chapter characterizes how legislators define the type of representation they provide to constituents and shows how this definition of representation changes in response to shifts in electoral pressure and changes in party control of Congress. To demonstrate this change, I use a large collection of every House press release from 2005 to 2010 and a statistical topic model that measures legislators' expressed priorities—their attention to salient topics. To increase substantive interpretability and address questions about the number of topics to include in the model, the model estimates a hierarchy of topics. A set of granular issue specific topics are nested in a set of coarse topics that capture broad differences in the content of press releases. Using estimates of legislators' attention to both types of topics, I show that, like senators, House members' expressed priorities lie on a credit claiming/position taking spectrum. And where House members fall on this spectrum depends not only on who they represent, but also responds to broad political changes. After the 2008 election, representatives' shift their expressed priorities—Republicans abandon credit claiming and articulate criticism towards Obama, while Democrats embrace credit claiming and defend the federal stimulus. Yet, even after responding to the changing conditions, legislators largely maintain the same broad style. The results in this chapter demonstrate the strategic ways legislators change how they communicate with constituents and demonstrates the utility of computational tools for studying representation.

1 Introduction

Communication is a central component of representation (Mansbridge, 2003; Disch, 2012). Legislators invest time and resources in crafting speeches in Congress, composing press releases to send to newspapers, and in distributing messages directly to their constituents (Yiannakis, 1982; Quinn et al., 2010; Lipinski, 2004; Grimmer, 2013). Indeed, the primary problem in studying the role of communication in representation is that legislators communicate so much that analysts are quickly overwhelmed. Traditional hand coding is simply unable to keep pace with the staggering amount of text that members of Congress produce each year.

In this chapter I use a *text as data* method and a collection of press releases to measure how legislators present their work to constituents (Grimmer and Stewart, 2013). Specifically, I measure legislators' expressed priorities: the attention they allocate to topics and issues when communicating with constituents (Grimmer, 2010). Using the measures of legislators' expressed priorities, I characterize how Republicans respond to the drastic change in institutional and electoral context after the 2008 election. Not only did the Republican party lose the White House, the Tea Party movement mobilized and articulated conservative objections to particularistic spending. Replicating a finding from Grimmer, Westwood and Messing (2014) with alternative measures, I show that Republicans abandon credit claiming. Instead, Republicans articulate criticisms of the Democratic party, the Obama administration, and Democratic policy proposals. In contrast, Democrats embrace credit claiming and defend Democratic policies—though less vocally than Republicans criticize those same proposals. In spite of the shifts in rhetoric, though, I demonstrate that there is a strong year-to-year relationship in legislators' presentational styles. So, while legislators are responsive at the margin to changing conditions, the basic strategy remains the same.

This chapter contributes to a growing literature that examines legislative speech using automated methods for text (Hillard, Purpura and Wilkerson, 2008; Monroe, Colaresi and Quinn, 2008; Quinn et al., 2010; Grimmer, 2013; Cormack, 2014). This literature has demonstrated how computational tools can be successfully used to examine the content of legislation and how the types of bills passed over time have change (Adler and Wilkerson, 2012). Other studies have demonstrated how text can be used to provide nuanced measures of legislators' ideal points (Gerrish and Blei, 2012). And still other studies have demonstrated how legislators use communication to create an impression of influence over expenditures (Grimmer, Westwood and Messing, 2014).

Like these prior studies, I exploit a large collection of Congressional text to study what legislators say and why it matters for representation. I use a collection of nearly 170,000 House press releases: every press release, from each House office, from 2005 to 2010. There is increasing evidence that press releases are a reliable and useful source for capturing how legislators communicate with their constituents. Grimmer (2013) shows that press releases contain politically relevant content not found in floor speeches and that press releases have a direct effect over the content of newspaper stories and constituent evaluations. This particular collection of press releases are also useful because they cover a tumultuous time in American history: including the wars in Iraq and Afghanistan, a financial and mortgage crisis that precipitated the deepest recession in a generation, and changes in party control of the Congress and presidency.

To examine the content of the press releases, I apply a model that estimates a hierarchy of topics and how legislators allocate their attention to each level of topics (Blaydes, Grimmer and McQueen, 2014). To construct the hierarchy of topics the model nests, or classifies, a set of granular topics into a set of coarse topics. This modeling strategy builds on Pachinko Allocation Models, that allow for a nesting of topics, while contributing a model that relies on a different distribution that allows for fast inferences (Li and McCallum, 2006). The model is useful both substantively and statistically. Substantively, the model provides an automatic classification between more position taking, credit claiming, and advertising press releases. Previous versions of topic models applied to Congressional communication required a second manual step to perform this classification (Quinn et al., 2010; Grimmer, 2013). This second step can be useful, but also can make analysis cumbersome and adds another layer of interpretation to the analysis.

Statistically, the model helps address concerns about selecting the number of topics in a model. One of the most consequential assumptions made when applying topic models is deciding how many topics to include in the analysis. Determining the number of topics is a particularly vexing problem for social scientists, because our goals when using unsupervised methods are often difficult to quantify (Grimmer and King, 2011) and because different types of analysis implies that different numbers of topics are ideal (Chang et al., 2009). The model in this chapter addresses this problem by providing two sets of topics. One set of topics are granular, or more specific, and are intended to capture legislators' attention to specific policy debates and actions that are discussed in the press releases. The second set of topics are coarse, or more broad, and capture broad differences in the types of language legislators use when communicating with constituents. By providing two types of topics, I show how the model facilitates an analysis of who discusses specific issues with constituents, while also facilitating broad comparisons in what legislators say to constituents.

Together, the model and data make possible new measures that help answer long standing questions about how political representation works in American politics (Mayhew, 1974; Fenno, 1978). As I discuss in the conclusion, this provides a demonstration of how large scale analysis of text can facilitate deeper and broader insights into how representation occurs in American politics.

2 Topic Models for Social Science

To analyze the collection of House press releases I use a topic model that estimates both coarse and granular topics. Topic models are an increasingly popular tool for studying large collections of texts (Blei, Ng and Jordan, 2003; Quinn et al., 2010). Topic models are an unsupervised tool that discovers the salient issues, or topics, in a collection of documents and then measures how attention to topics varies across documents, actors, or over time. Part of the reason for the popularity of topic models is that they exploit a hierarchical structure that is easily extended to include different features of the documents, the authors of the model, and when the documents were written (Blei and Lafferty, 2006; Quinn et al., 2010; Grimmer, 2010; Mimno and McCallum, 2008; Grimmer and Stewart, 2013). Exploiting this extensibility, Roberts, Stewart and Airoldi (2014) introduce the *Structural Topic Model* (STM): a general model that allows users to flexibly include a wide array of covariates to better understand how attention to topics varies and how different types of speakers discuss the same basic topic (see also Mimno and McCallum 2008).

Models like STM condition on a user provided set of characteristics. Other topic models, however, learn about groups from the analysis. For example, Grimmer (2013) introduces a model that groups legislators who dedicate similar attention to topics when communicating with constituents, while simultaneously estimating the topics of discussion and legislators' attention to those topics (see also Wallach 2008). The clustering of legislators has methodological benefits, by facilitating more accurate smoothing across individual senators. The clustering also provides substantive insights by creating coarse summaries of how legislators engage their constituents.

A closely related set of models group together topics that place similar emphasis on a the same set of words. Models such as Pachinko Allocation estimate a hierarchy of topics (Li and McCallum, 2006). At the top of the hierarchy are general topics that capture broad emphases in the texts. At the bottom of the hierarchy are more granular topics about narrower content in the documents. Like the clustering of authors based on their attention to topics, this grouping provides methodological advantages—ensuring that information is borrowed from topics that emphasize similar words.

The nesting of topics also helps address one of the major challenges in utilizing topic models in applied research. Like other unsupervised learning methods, topic models require users to set the number of topics that are used in the model. And determining how many components to include in a model remains one of the biggest challenges in applying topic models for social scientific research. Some methods attempt to avoid this assumption and use nonparametric priors to estimate the number of topics (Teh et al., 2006). But nonparametric priors are no panacea. Instead, models that make use of nonparametric priors substitute an explicit assumption about the number of topics to include in the model with an implicit assumption based on the properties of the particular nonparametric prior used (Wallach et al., 2010). This implicit assumption arises because nonparametric priors are not explicitly attempting to estimate the "correct" number of components to include in an unsupervised model, but instead they are attempting to estimate an underlying density. This is problematic, because Wallach et al. (2010) show that strong assumptions in the nonparametric priors determine the number of estimated components.

Alternatively, scholars have increasingly used task specific tests to determine the number of topics (Chang et al., 2009; Roberts et al., 2014*a*; Grimmer and Stewart, 2013). For example, it is common to select the number of topics that have the best predictive performance, but methods that perform well in prediction might have poor substantive properties (Chang et al., 2009). Other scholars have suggested methods that quantify the coherence of the topics, tying the evaluation closer to the way social scientists use topic models (Bischof and Airoldi, 2012; Roberts et al., 2014*a*; Grimmer, 2010).

These tests are useful, but still limit the application of any one instance of a topic model.

This limitation occurs because the optimal number of topics in any application depends on how the model and estimates from the model will be applied. When studying how legislators communicate with constituents, for example, more granular topics are useful when examining who participates in debates around policies, or for examining who claims credit for specific kinds of spending in districts. For other questions, however, a more coarse classification might be useful. For example, when making broad comparisons across legislators' styles, it may be useful to compare legislators' attention to credit claiming to their rates of position taking, regardless of what legislators claim credit for securing or what topics they articulate positions about.

Rather than estimate a single set of topics, this chapter uses a model first introduced in Blaydes, Grimmer and McQueen (2014) that estimates two different sets of topics, similar to the nesting of topics in Pachinko Allocation (Li and McCallum, 2006). The model that I use here has a two-layer hierarchy and nests granular topics into a set of coarse topics. The nesting allows us to naturally define subsets of topics that use broadly similar language, or language that accomplishes a similar substantive goal. While the nesting of topics actually increases the number of parameters to set when estimating the model, it also makes the final model fit more broadly applicable—ensuring the same model can be used to assess granular differences in the specific debates legislators participate in, while also making coarse comparisons across documents.

3 A Model for Nested Topics

To apply statistical models to the collection of press releases, I first preprocess the texts—representing its content as numbers. I do this using a standard set of techniques, though I slightly vary the recipe to account for idiosyncratic features of Congressional press releases (Grimmer and Stewart, 2013). I first make the most common, and perhaps most counterintuitive, assumption and discard word order (commonly referred to as the bag of words assumption). I also discarded punctuation and capitalization and stemmed the words, mapping words that refer to the same basic content to a common stem. I then removed words that occurred in less than 0.5% of the press releases, words that occurred in more than 90% of the press releases, stop words, and proper nouns that refer to specific Congressional districts, members of Congress, or American cities. Removing this set of words ensures that I do not obtain a set of region or Congressperson specific topics.

The result of the process is that for each legislator-year i ($i = 1, \ldots, 2,587$) I represent each press release j ($j = 1, \ldots, N_i$) as a $W = 2,727$ element-long count vector $\mathbf{y}_{ij} = (y_{ij1}, y_{ij2}, \ldots, y_{ij2727})$.[1] Each y_{ijw} counts the number of times token w occurs in document j from legislator i. Like Grimmer (2010), I model the collection of legislators' press releases as a mixture of von-Mises Fisher distributions, a distribution on a hypersphere: vectors that have (euclidean) length 1 (Banerjee et al., 2005; Grimmer, 2010; Gopal and Yang, 2014). To utilize the distribution, I work with a normalized version of the count vector, $\mathbf{y}^*_{ij} = \frac{\mathbf{y}_{ij}}{\sqrt{\mathbf{y}'_{ij}\mathbf{y}_{ij}}}$.

To construct the model, I suppose that for in each year, each representative in the House of Representatives, i, divides attention over a set of K topics $\boldsymbol{\pi}_i$, where π_{ik} represents the proportion of the representative's press release allocated to topic k. Throughout the analysis I will treat $\boldsymbol{\pi}_i$ as a measure of legislators' *expressed priorities*: the issues legislators emphasize when communicating with constituents. It is a priority because the model will not identify a particular policy position a legislator might take in their public statements. And it is expressed because the emphasis legislators give in their press releases to issues might differ from the time they spend working on those topics in the institution, or the legislators own personal priorities.[2]

The attention to topics is assumed to stochastically control the frequency of each topic

[1]There are 23 legislator years where I have no press releases from some legislators in a given year

[2]Grimmer (2013) shows that there is a correlation between how legislators behave in the institution and what the say to constituents.

in the collection of press releases. Each press release j from a legislator in a year is assumed to have one *granular* topic, which I represent with the indicator vector $\boldsymbol{\tau}_{ij}$. We assume that $\boldsymbol{\tau}_{ij} \sim \text{Multinomial}(1, \boldsymbol{\pi}_i)$. Given the granular topic, a document's content is drawn from a corresponding von Mises-Fisher distribution. That is, $\boldsymbol{y}^*_{ij}|t_{ijk} = 1 \sim \text{von Mises-Fisher}(\kappa, \boldsymbol{\mu}_k)$, where κ is a concentration parameter—analogous to the variance in a normal distribution—and $\boldsymbol{\mu}_k$ is a $2,727$ element long vector that describes the center of the topic .[3] If an entry of $\boldsymbol{\mu}_k$, μ_{kw} has a large weight, it implies that the token w is particularly prevalent in the topic.

To construct a hierarchy of topics I assume that the granular topics are nested in the coarse topics. Equivalently, the model simultaneously clusters documents into a set of granular topics and clusters granular topics into coarse topics. For each of the K granular topics I suppose that each granular topic belongs to one of C coarse topics. Let $\boldsymbol{\sigma}_k$ be an indicator vector for granular topic k: if $\sigma_{kc} = 1$ then granular topic k is assigned to coarse topic c. I suppose that $\sigma_{kc} \sim \text{Multinomial}(1, \boldsymbol{\beta})$ where $\boldsymbol{\beta}$ is a C element long vector that describes the proportion of granular topics assigned to each of the coarse topics. Given $\boldsymbol{\sigma}_k$, I then draw the granular topic from a von Mises-Fisher distribution with center at the corresponding coarse topic. Specifically, $\boldsymbol{\mu}_k|\boldsymbol{\sigma}_{kc} = 1 \sim \text{von Mises Fisher}(\kappa, \boldsymbol{\eta}_c)$. One of the virtues of using the von Mises-Fisher distribution is that it is conjugate to itself (Banerjee et al., 2005; Gopal and Yang, 2014), facilitating the hierarchy of topics.[4]

I follow Grimmer (2010) and set priors on $\boldsymbol{\pi}_i$, $\boldsymbol{\beta}$ and $\boldsymbol{\eta}_m$ to limit their influence on the parameters. The data generating process and priors implies the following hierarchical model,

[3]One might be concerned that the vMF distribution is inappropriate here, because of the zeros in the document. While this is a technical concern, in the actual application of the model the zeros, matter little, because there are many other plausible assumptions in the data generation process.

[4]Indeed, I could continue the hierarchy and create a nesting of the coarse topics. My experience has been, however, that in this setting another layer of topic clustering provides few insights and fairly noisy summaries of the texts.

$$\begin{aligned}
\boldsymbol{\pi}_i &\sim \text{Dirichlet}(0.01) \\
\boldsymbol{\eta}_c &\sim \text{vMF}(\kappa, \frac{\mathbf{1}}{\sqrt{\mathbf{2727}}}) \\
\boldsymbol{\beta} &\sim \text{Dirichlet}(1) \\
\boldsymbol{\sigma}_k &\sim \text{Multinomial}(1, \boldsymbol{\beta}) \\
\boldsymbol{\mu}_k | \sigma_{mk} = 1 &\sim \text{vMF}(\kappa, \boldsymbol{\eta}_m) \\
\boldsymbol{\tau}_{ij} | \boldsymbol{\pi}_i &\sim \text{Multinomial}(1, \boldsymbol{\pi}_i) \\
\boldsymbol{y}^*_{ij} | \tau_{ijk} = 1, \boldsymbol{\mu}_k &\sim \text{vMF}(\kappa, \mu_k)
\end{aligned}$$

which implies the following posterior distribution,

$$\begin{aligned}
p(\boldsymbol{\pi}, \boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\mu}, \boldsymbol{\tau} | \boldsymbol{Y}) \quad \propto \quad & \prod_{m=1}^{C} c(\kappa) \exp(\kappa \boldsymbol{\eta}_m^{'} \frac{\mathbf{1}}{\sqrt{\mathbf{2727}}}) \times \prod_{m=1}^{C} \prod_{k=1}^{K} \left[\beta_m c(\kappa) \exp(\kappa \boldsymbol{\mu}_k^{'} \boldsymbol{\eta_m})\right]^{\sigma_{m,k}} \times \\
& \prod_{i=1}^{2,727} \left[\prod_{k=1}^{K} \pi_{ik}^{-0.99} \times \prod_{j=1}^{N_i} \left[\pi_{ik} c(\kappa) \exp(\kappa \boldsymbol{y}^*_{ij} \boldsymbol{\mu}_k)\right]^{\tau_{ijk}} \right] \qquad (3.1)
\end{aligned}$$

where $c(\kappa)$ is a normalizing constant for the von Mises-Fisher distribution.

To approximate the posterior I use the variational approximation described in Blaydes, Grimmer and McQueen (2014).[5]

To apply this model (and other topics models), I have to assume the number of granular and coarse topics in the model. I select 44 granular topics—a number used in previous studies of Congressional communication—and 8 coarse topics. The number of coarse topics was determined after initial experiments with a subset of documents, but because the estimation is fully Bayesian, the model may do automatic model selection and select fewer topics. This occurs in this application, where only 7 of the coarse topics are assigned granular topics.

[5]Because I fix the prior on $\boldsymbol{\pi}_i$ we are able to avoid maximizing the dirichlet hyperparameters.

4 Validating the Topics and Legislators' Expressed Priorities

As Quinn et al. (2010) argue, unsupervised models require less work initially—estimating the topics of discussions—but then require more substantial investment to interpret their content. To begin interpreting the model output, Table 4 presents the coarse topics—between the horizontal lines—and the corresponding granular topics. The left-hand column contains a short description of each topic that I created after reading a random sample of press releases assigned to the category, the middle column provides words the 8 words that best distinguish the topic from the other the other topics, and the right-hand column presents the proportion of press releases that fall into the particular category.

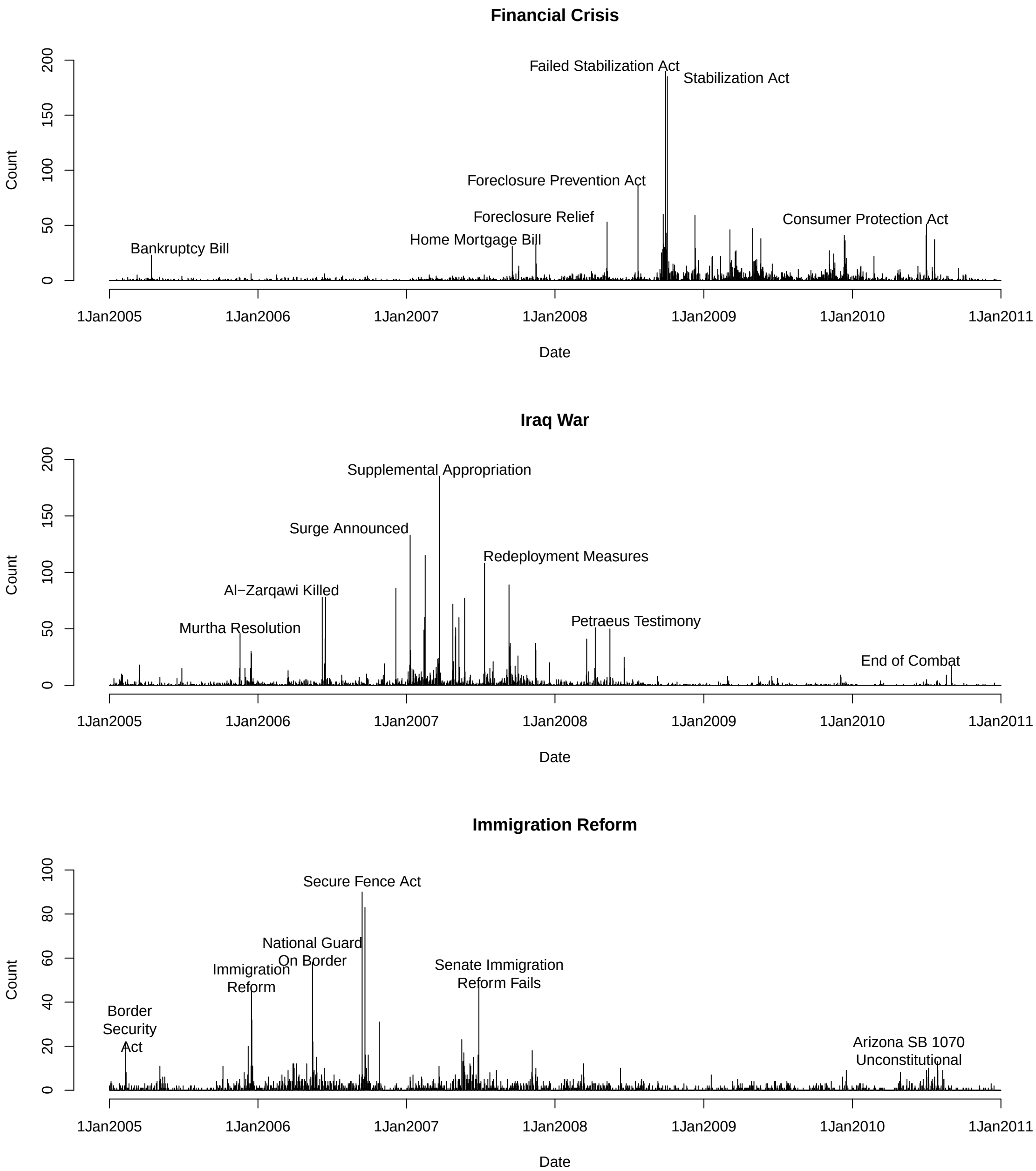
Pos. Taking/ Advertising	hous,tax,state,busi,vote,student,school,act	0.416
Committee Position	hous,committe,member,congress,repres,chairman,republican,democrat	0.046
Sponsored Leg.	act,legisl,law,protect,feder,hous,pass,introduc	0.043
International Disputes	state,unit,govern,israel,iran,intern,right,human	0.039
Taxes	tax,incom,relief,famili,credit,taxpay,increas,deduct	0.029
Health Leg.	health,care,insur,reform,cost,coverag,american,afford	0.028
Finance/Mortgage Crisis	financi,credit,mortgag,taxpay,consum,market,card,bank	0.027
Unemployment	job,unemploy,economi,econom,creat,american,stimulu,worker	0.026
Art Contests	school,student,high,art,district,competit,congression,educ	0.024
Office Hours/Internships	offic,district,congressman,constitu,staff,congression,hour,servic	0.023
Vote Explained	vote,right,congress,hous,elect,act,member,amend	0.02
Student Loans	student,educ,colleg,loan,program,school,higher,univers	0.02
Child. Issues	children,health,program,schip,care,insur,famili,child	0.017
Small Businesses	busi,small,job,tax,loan,sba,owner,econom	0.016
Prescription/Illicit Drugs	medicar,drug,senior,prescript,plan,enrol,beneficiari,benefit	0.016
Farming	farm,agricultur,farmer,program,produc,crop,food,usda	0.012
Trade	trade,agreement,china,worker,market,american,job,free	0.012
Service Academies	academi,nomin,student,school,militari,servic,high,appoint	0.011
Women's Issues	women,cancer,pay,equal,violenc,act,diseas,awar	0.008
Credit Claiming	fund,grant,water,program,feder,commun,project,airport	0.187
Stimulus Funding	fund,program,million,appropri,provid,billion,feder,help	0.035
Transportation Funds	project,fund,transport,million,improv,counti,feder,highway	0.034
Municipal Grants	commun,grant,develop,fund,counti,program,announc,rural	0.03
Fire Grants	grant,depart,firefight,program,equip,assist,announc,fund	0.025
FEMA/Disaster	disast,fema,assist,flood,feder,emerg,counti,hurrican	0.02
Health Spending	health,care,center,servic,medic,hospit,provid,commun	0.019
Water Grants/Resources	water,lake,project,river,great,resourc,fund,clean	0.017
Airport Grants	airport,grant,aviat,fund,faa,improv,runway,transport	0.007
Military Support/Budget	research,statement,budget,defens,militari,servic,famili,today	0.115
Statements	statement,today,datelin,fed,famili,peopl,death,pass	0.029
Veteran Service	servic,honor,militari,famili,serv,member,veteran,medal	0.027
Budget	budget,spend,fiscal,deficit,cut,tax,billion,year	0.025
Military Issues	defens,militari,forc,air,base,million,armi,nation	0.023
Research Support	research,cell,stem,scienc,univers,fund,diseas,develop	0.011
National Politics	presid,american,earmark,iraq,court,safeti,statement,congress	0.107
National Holidays	american,peopl,america,congress,nation,wage,work,day	0.027
Pres. Attack/Defend	presid,obama,bush,statement,address,american,union,congress	0.027
Iraq	iraq,troop,war,iraqi,presid,militari,forc,statement	0.025
Judiciary	court,suprem,right,decis,judg,rule,law,justic	0.013
Cons./Employee Safety	safeti,food,product,fda,consum,drug,recal,children	0.01
Earmarks	earmark,egregi,week,spend,republican,project,appropri,reform	0.005
District Positions	veteran,energi,new,oil,va,price,fuel,increas	0.091
Oil and Gas	energi,oil,price,fuel,renew,effici,increas,product	0.042
Veterans Care	veteran,va,care,affair,servic,benefit,militari,medic	0.028
Ribbon Cutting	new,citi,state,facil,said,site,region,area	0.021
National Security	secur,nation,border,immigr,illeg,homeland,social,law	0.068
National Park	nation,park,guard,histor,land,area,forest,preserv	0.025
Homeland Security	secur,social,homeland,port,terrorist,nation,attack,protect	0.024
Immigration	immigr,border,illeg,secur,law,enforc,alien,reform	0.02
District Meetings	meet,hall,town,counti,constitu,congressman,district,street	0.016
District Meetings	meet,hall,town,counti,congressman,constitu,district,street	0.016

Table 4 demonstrates that the model is able to both identify distinct granular topics in the press releases and that the coarse topics identify substantively interesting groups of press releases. Consider, for example, the *Credit Claiming* coarse topic, which identifies press releases legislators use to receive credit for expenditures that occur in their district (Mayhew, 1974; Grimmer, Westwood and Messing, 2014). The granular topics assigned to the credit claiming coarse topic each claim credit for different kinds of expenditures. For example, Stephanie Herseth (D-S.D.) (later Herseth-Sandlin) issued a press release that "announced $3 million in appropriations funding for a new water well at Ellsworth Air Force Base" (Herseth, 2006) and Rep. Rodney Alexander (R-LA) "and Sen. David Vitter announced today that Evangeline Parish will receive a federal grant in the amount of $74,980 to purchase and install equipment to improve the water system " (Alexander, 2007). Other legislators claim credit for grants to airports, such as Bart Stupak who "announced three airports in northernMichigan have received grants totaling $726,409 for airport maintenance and improvements" (Stupak, 2010). And other legislators claim credit for grants to fire departments in their district, such as Brian Higgins (D-NY) who "announced Bemus Point Volunteer Fire Department will receive $43,966 in federal Homeland Security funding" (Higgins, 2006). While the legislators are claiming credit for different types of expenditures, they are engaging in the same activity: ensuring they receive credit for spending in the district. To do this, the legislators use distinct language that the coarse topic identifies—*announcing funds* for *projects* in their district. The hierarchical model, then, is able to identify a category of political action that previous qualitative scholarship had identified (Mayhew, 1974; Fenno, 1978) and other applications of topic models had to manually categorize topics after the model was run (Grimmer, 2013). Other coarse topics identify distinct ways that legislators discuss their work with constituents. The most prevalent coarse topic identifies positions legislators take, positions they hold in Washington, or services that they perform for constituents. Other coarse topics identify debates about national politics, support for the military, and national security. Rather than requiring an ad hoc second step or manual labeling, then, the coarse topics identify substantively interesting groups of topics automatically from the collection of press releases.

Within the coarse topics the granular topics identify areas of salient policy disputes. For example, the granular *Iraq* topic, nested in the *National Politics*, coarse topic, identifies press releases about the second Gulf war. In 2007 Michael Capuano, a liberal Democrat from Massachusetts, explained that he "pushed for a vote on a course of action that would have gotten us out of Iraq much sooner and stipulated that all funding go toward drawing down troops" (Capuano, 2007), while Jerry Lewis, a more conservative Republican from California, criticized a supplemental spending bill for the war, arguing that "this legislation does not accurately reflect the will of the American public...but rather the desires of Speaker Pelosi and the Abandon Our Troops Caucus within the Democratic Party" (Lewis, 2007). Other topics identify press releases about a wide range of substantive topics—such as the financial crisis, rising unemployment, farming, and immigration. Other topics discuss ways legislators directly engage constituents—including district meetings, Congressional art contents, service academy nominations, and internships in Congressional offices.

While validating each of the individual topics is infeasible for this single chapter, I can examine over time variation in the prevalence of topics as a measure of face validity of the topics (Quinn et al., 2010). Figure 3 shows the daily count of press releases from the financial crisis (top plot), the Iraq War (middle plot), and Immigration Reform (bottom plot). Each plot shows that spikes in attention to each topic corresponds with major events. For example, the days with the most press releases about the financial crisis correspond with the Congressional debate and initial inaction at the height of the financial crisis. There are similar spikes in attention to the Iraq War as legislators debated supplemental spending bills that redeployed US troops and a spike at the end of combat in Iraq. The large increases in attention that correspond with actual events are evidence that the granular topics are valid—estimating the content I claim they are estimating.

Figure 1: Spikes in Topics Correspond with Real World Events

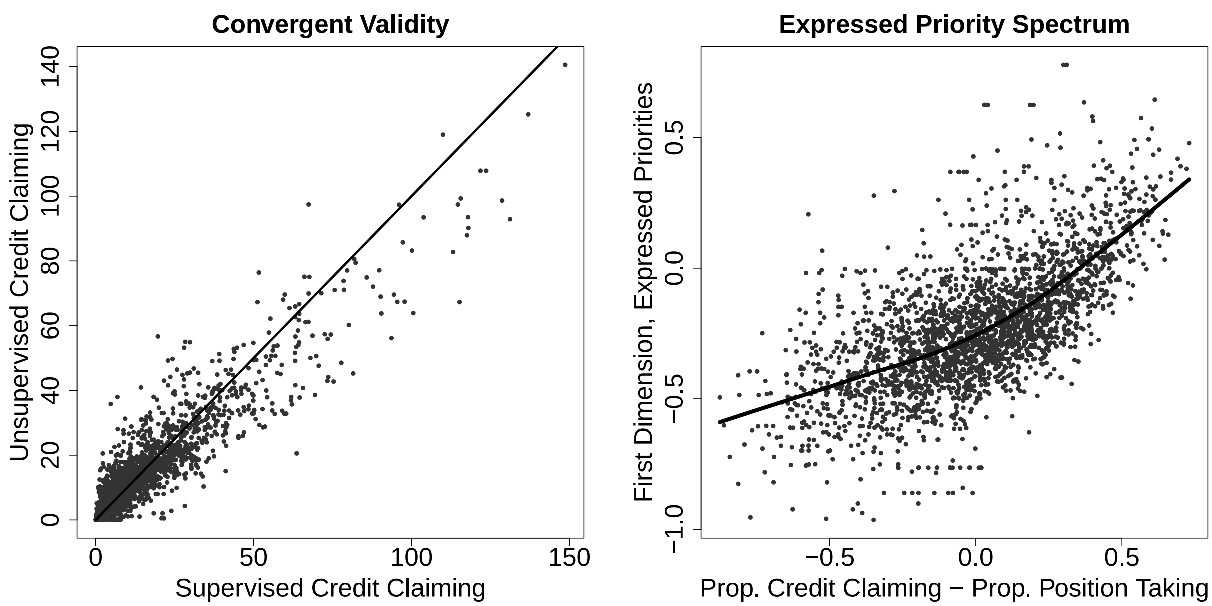


This figure shows that large spikes in attention to topics corresponds with salient events that drive Congressional attention. This is evidence that the granular topics are valid.

Not only do I have to demonstrate that the estimated topics are valid—capturing the types of rhetoric I claim that they are—I also have to demonstrate that my measures of legislators' expressed priorities accurately capture how legislators explain their work in Washington to constituents. Figure 2 presents evidence that the model is accurately estimating how legislators divide their attention in press releases. As Grimmer and Stewart (2013) argue, one of the most stringent convergent validity checks for an unsupervised learning method is to compare its output to estimates from hand coded documents that are intended to measure the same topic or concept. The left-hand plot in Figure 2 carries out this comparison for a subset of the model. The vertical axis in Figure 2 shows the estimated number of credit claiming press releases from each legislator in each year from the unsupervised model used in this chapter. The horizontal axis shows the estimated number of credit claiming press releases from a model explicitly designed to identify credit claiming press releases. Specifically, the estimates are from Grimmer, Westwood and Messing (2014), who used a team of well trained coders to hand code 800 press releases as credit claiming or not. Then, Grimmer, Westwood and Messing (2014) used an ensemble of supervised learning techniques to classify the remaining press releases. The black line is a 45 degree line, where the points would align if there is a perfect relationship between the two measures. The tight clustering of points along the line provide visual evidence for the strong correlation of 0.93. And the strong correlation remains if I compare the proportion of press releases allocated to credit claiming (0.79).

The strong correlation between supervised and unsupervised credit claiming provides evidence that our model is accurately estimating how legislators portray themselves to constituents. As a second validity check I can examine whether expected variation in legislators' behavior manifests in our measures. A reasonable expectation is that legislators will discuss prominent industries in their district more often (Adler and Lapinski, 1997). As a simple assessment of this expectation, I regressed the proportion of press releases in the farming

Figure 2: Expressed Priority Measures Converge with Previous Measures and Display Similar Variation



The left-hand plot shows that the measure of credit claiming from the unsupervised model is closely related to measures of credit claiming from a supervised model, as presented in Grimmer, Westwood and Messing (2014). The right-hand plot shows that the primary variation underlying legislators' expressed priorities is a position taking-credit claiming spectrum.

category on the proportion of employed constituents who work in farming. This reveals that legislators who represent farming districts discuss agriculture more often. Indeed, representatives from districts at the 90th percentile of farm employment discuss allocate 1.4 percentage points more of their press releases to agriculture than colleagues with few farm jobs in the district (95 percent confidence interval [0.01, 0.02]).

Legislators' position in the institution is another likely predictor of how they present their work to constituents. For example, one might expect that members of the Appropriations Committee focus more on credit claiming than other legislators. And the measures from the model suggest they do. Members of the Appropriations Committee allocate 5.9 percentage points more of their press releases to credit claiming than other representatives (95 percent confidence interval, [0.04, 0.07]). Similarly, one might expect that Congressional leaders will focus less on claiming credit for expenditures and focus more of their attention on broad national issues—an expectation that manifests in our measured credit claiming. Party leaders in the House allocate 4.0 percentage points fewer of their press releases to credit claiming (95 percent confidence interval [-0.07,-0.01]) and 6.2 percentage points more to the national issues topic (95 percent confidence interval [0.04, 0.08])

My estimates of House members' presentational styles also exhibit variation that is consistent with other well validated measures of other representatives' presentational styles. Grimmer (2013) shows that legislators' expressed priorities lies on a credit claiming, position taking spectrum. The right-hand plot in Figure 2 shows this same variation underlies House members' press releases. The right-hand figure plots the principal component of legislators' expressed priorities against the proportion of press releases claim credit, less the proportion of press releases where legislators articulate positions. There is a strong correlation between the two measures of 0.65, and a correlation of 0.71 between legislators' credit claiming rates and the principal component underlying the expressed priorities. Like senators, House members must decide how to balance claiming credit for spending that happens in their district.

And like senators, who House members represent is correlated with how they balance credit claiming and position taking. Grimmer (2013) demonstrates that senators who misaligned with their districts focus more on credit claiming, while aligned legislators articulate more positions. House members from marginal districts allocate 5.4 percentage points more of their press releases to credit claiming than their more aligned colleagues (95 percent confidence interval [0.04, 0.07]), while more aligned representatives 5.0 percentage points more of their press releases to national topics (95 percent confidence interval [0.04, 0.06]) and allocate 3.8 percentage points more likely to articulating positions and advertising (95 percent confidence [0.03, 0.05]).

The unsupervised model is able to accurately identify both coarse and granular topics and to reliably estimate legislators' credit claiming propensity. In the next section I use the estimated priorities to show how members of the Republican party shift their attention after Barack Obama's election toward criticism and away from claiming credit for expenditures in their district.

5 Stability and Change in Legislators' Expressed Priorities

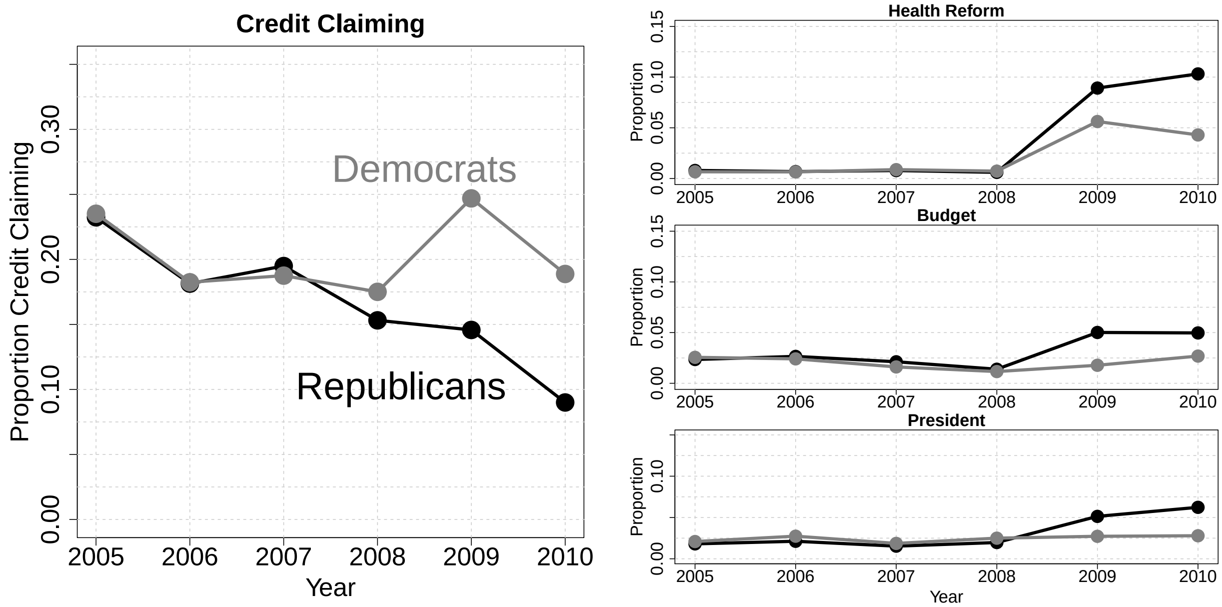
The collection of press releases used in this chapter covers six volatile years in American history and in Congress, reflecting change in who held institutional power in Congress and the electoral pressure representatives—particularly Republicans—felt from their base. Republicans held the House majority in 2005, but lost that majority in the 2006 elections. The Republicans were then routed in both the 2008 Congressional and presidential elections. Not only did they lose the White House, they surrendered more seats in both the House and Senate, bolstering the Democratic party's majority. And as the newly elected Congress and

Obama passed stimulus measures and began considering health reform a mass movement of conservatives—the "Tea Party"—articulated frustrations with stimulus spending, health care reform, and the Obama administration policies (Skocpol and Williamson, 2011). The conservative movement pressured Republican members of Congress to reject particularistic expenditures and to lower taxes and threatened those who failed to change with primary challenge.

In response to the changing electoral and institutional pressures, Republicans altered how they presented their work to constituents. The left-hand plot in Figure 3 replicates a plot from Grimmer, Westwood and Messing (2014), but uses the unsupervised measures of credit claiming to show that after Obama's election Republicans allocated a much smaller percentage of their press releases to credit claiming than in previous years. In 2005 Republicans legislators claimed credit for spending in 23.2% of their press releases, while Democrats claimed credit in 23.5% of theirs. By 2010, however, Republicans allocated only 9% of their press releases to credit claiming—a 5.5 percentage point decline from 2010. The shift in Republican credit claiming is primarily due to conversion: from 2009 to 2010 Republicans decreased their credit claiming rate 6.7 percentage points (95 percent confidence interval [-0.09, -0.05]). In contrast to the Republican aversion to credit claiming, Democratic credit claiming spiked in 2009, when Democrats allocate 24.7 percent of their press releases to claiming credit for spending.

Instead of claiming credit, Republicans amplified criticism of the Obama administration and Democratic policies. The top plot in the right-hand side of Figure 3 shows the proportion of press releases Republicans (black line) and Democrats (gray line) allocated to health care reform. As the legislation that would eventually become the Affordable Care Act worked through Congress, both Democrats and Republicans increased the frequency of press releases about health care reform, but Republican House members were especially focused on health care reform. In 2009 Republican House members allocated 3.3 percentage points more of

Figure 3: Republicans Avoid Credit Claiming and Instead Attack Presidential Policies



Republicans avoid claiming credit for spending and instead focus on criticism Obama administration policies

their press releases to health reform than Democrats (95 percent confidence interval [0.02, 0.05]) and this difference grew to 6.0 percentage points in 2010 (95 percent confidence interval [0.04, 0.08]).

When Republicans discussed the health care reform they were critical of the content of the legislation and the legislative procedure to pass it. For example, Adam Putnam (R-FL) expressed skepticism about the potential benefits of health care reform, because "despite the president's very rosy view of cost savings, I think most Americans have learned through hard experience to be skeptical of such claims" (Putnam, 2009). Ralph Hall (R-TX) offered a similar condemnation of the legislation, arguing that "We need health care reform, but the Democrats' radical plan is not the prescription for reform that the American people want or deserve" (Hall, 2010). Republicans also expressed dismay that Democrats decided "to break their promise of open and informed debate over Health Care" (Linder, 2010) and warned their constituents that "In Washington, we're witnessing...Pelosi Madness...as the

Speaker attempts to push through this health care legislation, regardless of cost, the desire of the American people and transparency." (Sensenbrenner, 2010). Even after the legislation was passed, Republicans like Tom Price (R-GA) criticized the law, arguing that "Democrats ignored the Constitution in order to pass a law that would put Washington in control of your personal health care, while curtailing access to quality, affordable health care" (Price, 2010).

The growing Tea Party movement also conveyed dismay at particularistic expenditures (Skocpol and Williamson, 2011). The middle right hand plot in Figure 3 shows that, consistent with the Tea Party rhetoric, Republicans also attacked Obama and Democrats on particularistic expenditures in the stimulus and spending more generally. In 2008, both Republicans and Democrats allocated about the same attention to discussing the federal budget: Republicans allocated 1.4% of their press releases to budget issues, only slightly more than the 1.2% of press releases Democrats allocated to the budget in their press releases. By 2009, however, a large difference emerged: Republicans allocated 5.0% of their press releases to the budget, a 3.6 percentage point increase from 2008 and 3.2 percentage points more than Democratic House members. And Republicans maintained their focus on the budget in 2010, allocating 5.1% of their press releases to the budget and spending issues.

Just like the health care debates, Republicans are sharply critical of the Obama administration when discussing the budget. For example, Todd Akin (R-MO) argued that the American Recovery and Reinvestment Act "places an additional $800 Billion on top of historic levels of debt, and without the realistic promise of actual job creation" (Akin, 2009). Eric Cantor (R-VA) criticized a budget proposal because " The President's budget spends more than any other in history, creates the largest deficits in history, and imposes the largest tax increases in history - at a time when our country can least afford it" (Cantor, 2010). And Mary Bono-Mack alleged that "The passage of the state bailout bill is yet another example of the Democrats' tax and spend policies which are compromising our nation's future and the futures of our children and grandchildren" (Bono-Mack, 2010). The differences are evident in more quantitative comparisons of Republican and Democratic language when discussing the stimulus. Republicans use words like *spend*, *govern*, *democrat*, *taxpayer*, and *trillion* at a much higher rate than Democrats, who use words like *budget*, *cut*, and *education* more often than Republicans. The bottom plot in Figure 3 shows a similar increase in Republican criticism of the president.

The shift in rhetoric is evidence that legislators' expressed priorities are responsive to changing conditions in districts and oppositions. Yet, there remains stability in legislators' expressed priorities from year to year. One way to measure this is to assess the correlation across years in the proportion of press releases legislators allocate to the coarse topics, given that legislators remain in Congress. Overall, there is a year-to-year correlation of 0.81 in legislators' expressed priorities. This strong correlation is found in credit claiming (0.60) and is particularly strong in non-credit claiming coarse topics (0.83). The stability is not just found in the year-to-year measurse of legislators' expressed priorities. The strong correlation is even present over the entire six year period studied here. The correlation between House members' expressed priorities in 2005 and expressed priorities in 2010 is 0.72.

Even though Republicans and Democrats shift their rhetoric as different policy proposals are considered or in response to pressure from the base, legislators maintain largely the same style over the 6 years. This provides insight into the origins of legislators' presentational styles. Using one of the most volatile time periods in recent political history, we see evidence that legislators adjust how they discuss their work with constituents in response to changing electoral and institutional conditions. But the response is on the margin and a deviation from a longer run strategy that legislators develop over the course of their career (Grimmer, 2013). And as a result, there remains a strong over time relationship in legislators' expressed priorities.

6 Conclusion

A growing literature shows how legislators use communication to shape their relationship with constituents. This chapter contributes to this literature, providing new measures of how House members' expressed priorities respond to tumultuous changes in institutional and electoral contexts. To measure the expressed priorities, I use a large collection of House press releases and a statistical topic model that estimates granular and coarse topics, along with legislators' attention to those topics. The model provides two different types of topics, facilitating granular analysis of legislators' attention to more specific policy areas and more general behavior, such as credit claiming.

Using the measures from the model, I show Republican House members abandon credit claiming after Obama's election, while Democratic House members amplify their credit claiming. In place of claiming credit for money, Republican House members criticize the Affordable Care Act, stimulus expenditures, and more generally the Obama administration. Even though there is a shift in rhetoric after the 2008 elections, I show that legislators' attention to the coarse topics are broadly stable over time. This demonstrates legislators' ability to respond to changing institutional and electoral conditions, but this is a change on the margin, from legislators' persistent strategies.

Computational tools make possible studying how legislators directly engage their constituents and how this engagement matters for representation. And models that are developed to better understand how Congress works can contribute to many other substantive areas. Consider, for example, how we understand the relationship between legislators' work in Washington and what they say about that work to constituents. At the moment measures are developed in individual areas and related to each other in a second stage analysis. A more productive model might link the activities in a single model, allowing a creation of a comprehensive measure of how legislators approach diverse areas of their work (Bernhard, Sulkin and Sewell, 2014). Such a model would be useful in any setting where scholars wanted to link text with non-text data to facilitate inferences.

Computational tools could also be useful in understanding the effects of legislators' statements on constituents. For example Grimmer, Westwood and Messing (2014) use text analysis tools to motivate experiments that demonstrate the effect of legislators' messages on constituents. To do this Grimmer, Westwood and Messing (2014) had to determine the most salient features of the credit claiming messages to vary. But pairing randomly assignment with machine learning methods could facilitate *discovery* of the features of messages that are likely to have the largest effect on constituent response. In general, there is a need to better understand how to understand causal inference and text analysis methods (Roberts et al., 2014*b*).

Expanded resources and models also facilitate inferences that were previously impossible. For example, previous work has analyzed how legislators are covered in local papers and how legislators' own efforts to alter how they are covered in papers (Arnold, 2004). Yet, technological limitations limit the scope of what we can learn about how legislators are covered in prior work. Computational tools and digital collections of news, however, facilitate insights into how legislators are covered across diverse outlets and over extended time periods. Likewise, we know little about what constituents say when communicating with their legislators (Butler, 2014; Grose, Malhotra and Van Houweling, 2014). And while it is unlikely Congressional offices will provide access to letters from constituents, social media provides an opportunity to study how the public pressures representatives in public settings.

A common theme in this future work is that a combination of new digital records of text and statistical tools for analyzing the large collections will provide deep insights into how representation occurs in American politics.

References

Adler, E. Scott and John S. Lapinski. 1997. "Demand-Side Theory and Congressional Committee Composition: A Constituency Characteristics Approach." *American Journal of Political Science* 41(3):895–918.

Adler, E. Scott and John Wilkerson. 2012. *Congress and the Politics of Problem Solving.* Cambridge University Press.

Akin, Todd. 2009. "Rep. Akin Opposes Big Government Spending Plan." Representative Press Release.

Alexander, Rodney. 2007. "Rep. Alexander, Sen. Vitter Announce Grant Funding for Evangeline Parish." Representative Press Release.

Arnold, R. Douglas. 2004. *Congress, the Press, and Political Accountability.* Princeton Press.

Banerjee, Arindam, Inderjit S Dhillon, Joydeep Ghosh and Suvrit Sra. 2005. "Clustering on the Unit Hypersphere using von Mises-Fisher Distributions." *Journal of Machine Learning Research* 6:1345–1382.

Bernhard, William, Tracy Sulkin and Daniel Sewell. 2014. "Legislative Style." University of Illinois Mimeo.

Bischof, Jonathan and Edoardo Airoldi. 2012. Summarizing Topical Content with Word Frequency and Exclusivity. In *International Conference on Machine Learning*, ed. John Langford and Joelle Pinelle. pp. 201–208.

Blaydes, Lisa, Justin Grimmer and Alison McQueen. 2014. "Mirrors for Princes and Sultans: Advice on the Art of Governance in the Medieval Christian and Islamic Worlds." Stanford University Mimeo.

Blei, David, Andrew Ng and Michael Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning and Research* 3:993–1022.

Blei, David and John Lafferty. 2006. "Dynamic Topic Models." *Proceedings of the 23rd International Conference on Machine Learning* 23.

Bono-Mack, Mary. 2010. "Congress Needs to End Its Tax Spend Policies." Representative Press Release.

Butler, Daniel. 2014. "The Discriminating Politician." Washington University St. Louis Mimeo.

Cantor, Eric. 2010. "Rep. Cantor on Fiscal Year 2011 Budget." Representative Press Release.

Capuano, Michael. 2007. "Rep. Capuano Issues Statement on Iraq, H.R. 1591: The Iraq Accountability Act of 2007." Representative Press Release.

Chang, Jonathan, Jordan Boyd-Graber, Sean Gerrish, Chong Wang and David Blei. 2009. Reading Tea Leaves: How Humans Interpret Topic Models. In *Neural Information Processing Systems*.

Cormack, Lindsey. 2014. "Sins of Omission: Legislator (Mis)Representation in Constituent Communications." New York University Mimeo.

Disch, Lisa. 2012. "Democratic Representation and the Constituency Paradox." *Perspectives on Politics* 10(3):599–616.

Fenno, Richard. 1978. *Home Style: House Members in their Districts.* Addison Wesley.

Gerrish, Sean and David Blei. 2012. "How They Vote: Issue-Adjusted Models of Legislative Behavior." Neural Information Processing Systems.

Gopal, Siddharth and Yiming Yang. 2014. "Von Mises-Fisher Clustering Models." *Journal of Machine Learning Research* 32(1):154–162.

Grimmer, Justin. 2010. "A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases." *Political Analysis* 18(1):1–35.

Grimmer, Justin. 2013. *Representational Style: What Legislators Say and Why It Matters.* Cambridge University Press.

Grimmer, Justin and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* . Forthcoming.

Grimmer, Justin and Gary King. 2011. "General Purpose Computer-Assisted Clustering and Conceptualization." *Proceedings of the National Academy of Sciences* 108(7):2643–2650.

Grimmer, Justin, Sean J. Westwood and Solomon Messing. 2014. *The Impression of Influence: Legislator Communication, Representation, and Democratic Accountability*. Princeton University Press.

Grose, Christian, Neil Malhotra and Robert P. Van Houweling. 2014. "Explaining Explanations: How Legislators Explain their Policy Position and How Citizens React." Stanford University Mimeo.

Hall, Ralph. 2010. "Rep. Hall Urges Democrats to Drop Plans for Government Takeover of Health Care." Representative Press Release.

Herseth, Stephanie. 2006. "Rep. Herseth Announces $3 Million for Ellsworth Water Well." Representative Press Release.

Higgins, Brian. 2006. "Higgins Announces Federal Grant for Bemus Point Fire Department." Representative Press Release.

Hillard, Dustin, Stephen Purpura and John Wilkerson. 2008. "Computer-Assisted Topic Classification for Mixed-Methods Social Science Research." *Journal of Information Technology & Politics* 4(4):31–46.

Lewis, Jerry. 2007. "Rep. Lewis Denounces Supplemental Spending Bill That Let's Politicians Micromanage Troop Decisions." Representative Press Release.

Li, W. and A. McCallum. 2006. Pachinko Allocation: DAG-Structure Mixture Models of Topic Correlations. In *International Conference on Machine Learning*. pp. 577–584.

Linder, John. 2010. "Linder Statement on C-Span Call for Transparency." Representative Press Release.

Lipinski, Daniel. 2004. *Congressional Communication: Content and Consequences*. University of Michigan Press.

Mansbridge, Jane. 2003. "Rethinking Representation." *American Political Science Review* 97(4):515–528.

Mayhew, David. 1974. *Congress: The Electoral Connection*. Yale University Press.

Mimno, David and Andrew McCallum. 2008. "Topic Models Conditioned on Arbitrary Features with Dirichlet-multinomial Regression." *Conference on Uncertainty in Artificial Intelligence* . Plenary Presentation.

Monroe, Burt, Michael Colaresi and Kevin Quinn. 2008. "Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict." *Political Analysis* 16(4):372.

Price, Tom. 2010. "Rep. Price Issues Statement on Court Ruling that Obamacare is Unconstitutional." Representative Press Release.

Putnam, Adam. 2009. "Rep. Putnam Issues Statement on President Obama's Health Care Speech." Representative Press Release.

Quinn, Kevin et al. 2010. "How to Analyze Political Attention with Minimal Assumptions and Costs." *American Journal of Political Science* 54(1).

Roberts, Margaret E., Brandon M. Stewart and Edoardo M. Airoldi. 2014. "Structural Topic Models." Harvard University Mimeo.

Roberts, Margaret E. et al. 2014*a*. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* . Forthcoming.

Roberts, Margaret E et al. 2014*b*. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58(4):1064–1082.

Sensenbrenner, James. 2010. "March Madness...Washington Style." Representative Press Release.

Skocpol, Theda and Vanessa Williamson. 2011. *The Tea Party and the Remaking of Republican Conservatism.* Oxford University Press.

Stupak, Bart. 2010. "Rep. Stupak Announces $726,409 for Airports in Alpena, Delta, Chippewa Counties." Represenative Press Release.

Teh, Yee Weh, Michael Jordan, Matthew Beal and David Blei. 2006. "Hierarchical Dirichlet Processes." *Journal of the American Statistical Association* 101(476):1566–1581.

Wallach, Hanna. 2008. Structured Topic Models for Language PhD thesis University of Cambridge.

Wallach, Hanna, Lee Dicker, Shane Jensen and Katherine Heller. 2010. An Alternative Prior for Nonparametric Bayesian Clustering. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS).* Vol. 9.

Yiannakis, Diana Evans. 1982. “House Members’ Communication Styles: Newsletter and Press Releases.” *The Journal of Politics* 44(4):1049–1071.