Congressional Communication: The Ideological Nature of Official E-mail

Lindsey Cormack
Department of Politics
New York University
19 West 4th Street, 2nd Floor
New York, NY 10012-1119

October 25, 2013

Abstract

This paper analyzes the e-mail messages that members of Congress send to constituents. Using a newly collected dataset, I test whether legislators can be ordered ideologically using only the content of these messages. If legislators communicate predominantly with policy-devoid “advertising” messages and avoid political “position taking” in the manner expected by Mayhew (1974), then their messages should not be very ideologically informative. The results presented here show that legislators can be ordered using only the language sent in messages to constituents, suggesting that position taking is a more often used communication strategy than advertising in the modern Congress.


1 Introduction

At the time House members are called on to vote they are very aware that they may be called on to explain their vote to some of their constituents. The anticipated need to explain influences their decision on how to vote. They may cast a certain vote only when they are convinced that they have a satisfactory explanation in hand. To them, therefore, voting and explaining are interrelated aspects of a single strategic problem. If that is the way members of Congress see it, then it might be useful for political scientists to look at it that way as well - to spend a little less of our time explaining votes and a little more time explaining explanations. - Home Style, Fenno 1978

Our knowledge of the general ideological contours and attendant polarization of Congress is largely drawn from voting records and various vote scaling techniques. In this paper I add to this work by analyzing the e-mail messages that members of Congress send to their constituents in order to better understand the nuance of how ideological position taking may be communicated to constituents. I test whether legislators can be ordered ideologically using only the texts of these messages. The results of my analyses urge a reexamination of the dominant theory characterizing Congressional communications, namely that legislators will mostly send policy-devoid “advertising” messages and avoid political “position taking” as their dominant strategy (Mayhew 1974).

At the forefront of most theories of legislative behavior is Mayhew’s (1974) argument that a legislator’s preeminent goal is reelection. Fenno (1978) hints at the same driving force for legislators, by arguing that legislators develop home styles and communication techniques in a manner that optimizes their chances of reelection. That is, communicating to one’s constituents by describing what happens during the act of representation is key to being reelected. To that end, I ask the following questions: Does a liberal legislator communicate to her constituents in a way that is different than a conservative legislator? If so, can these differences be generalized across an ideological continuum?

If the messages sent by legislators are ideologically organized in a way that generally
corresponds to our placement of legislators estimated from their voting records, then the messages legislators send could actively contribute to a greater feeling of polarization in the electorate. Assuming political preferences are dependent on socially supplied information, as Huckfeldt and Sprague (1987) argue, the ideological content of legislator e-mails are doubly important because they might shape constituent views, rather than just reflecting them. If constituents read e-newsletters that are politicized and partisan-polarized, constituent views of, and demands on politicians may in turn become even more polarizing. This is similar to the logic Mayhew (1974) uses to describe why the incumbency advantage is increasing over time; as a legislator presents himself as an “errand boy” that helps constituents wade through bureaucracy, constituents in turn prefer the senior (incumbent) members of Congress who have more experience in dealing with the bureaucracy.

In the following sections, I first discuss competing theories about why a legislator might communicate ideologically. Second, I review the modern communication techniques legislators have at their disposal. Then, I introduce a new dataset of official legislator e-newsletters during the 111th Congress. Next, I present an analysis of these official communications along with a series of robustness tests. Lastly, I conclude that, on average, legislators communicate ideologically and discuss implications from this finding.

2 Theoretical Directions

Work on political communication has a long, but somewhat scattered history. Research has largely been impressionistic and anecdotal rather than systematic. This characterization of the literature reflects the difficulties of recording natural language texts from a body of over 500 members. Fenno was one of the first in the field to record the art of “presentation of self” that legislators use when they communicate with their constituents in his 1978 book, Home Style. A more recent attempt to capture the communication behaviors of legislators is Lip-
inski’s 2004 book, *Congressional Communication: Content and Consequences* that analyzes franked messages from 100 randomly selected members of the House. The most current work in this tradition is Grimmer’s forthcoming book, *Representational Style: What Legislators Say and Why It Matters* (forthcoming). In his research, Grimmer analyzes Senator press releases and credit claiming behavior. Beyond these 3 major works, the field is sprinkled with other insightful pieces, but no work has yet examined legislator to constituent messages sent over the medium of e-mail.

Because of the difficult nature of these wide spanning studies, theories about the ideological content of legislator to constituent communications have been largely speculative and rarely evaluated quantitatively. While Fenno reports that different legislators adopt different home styles, his sample size and data recording techniques do not allow him to test for variations in communicative behavior across ideological lines. Lipinski, while having more data than Fenno, still suffers from a medium (franked paper mail) that does not readily lend itself to the type of analyses that can be done on electronic texts; he was also forced to rely on a random sample of 100 House members because of all the manual classifications involved in his work. Grimmer has electronically recorded press releases of Senators, and so the type of the analysis I perform on legislator e-mails is certainly possible, although he does not pursue this. Importantly, press releases are different from e-mails because they are typically written for newspapers to reprint, whereas e-newsletters are constructed to be immediately delivered to a constituent inbox; therefore the expectations of the types of messages sent are likely to differ.

These works, along with others from different political science literatures, inform my expectation regarding the ideological content of legislator e-mails. The general test is seemingly simple, either legislators communicate in a way that is ideologically organized or they do not. I discuss the theoretical arguments that can be marshaled to support both sides of this supposition in the following two sections.
2.1 Non-ideological Communication

Mayhew (1974) presents three types of legislator communication strategies: advertising, credit claiming, and position taking. He argues that all legislators have an incentive to advertise their name. Advertising is, “any effort to disseminate ones name among constituents in such a fashion as to create a favorable image but in messages having little or no issue content.” Credit claiming involves giving the impression of being pivotal for some particularistic benefit for the home district. Position taking is the most political of strategies because when position taking, a legislator must support or oppose some sort of policy action or idea. Out of all three strategies, advertising is the easiest in the sense that simply giving basic information associated with one’s name and office is more straightforward than describing how a legislator is responsible for a specific governmental action or resource allocation, and runs no risk of alienating potential voters in a manner that position taking can.

In one of the best known works on legislator behavior in constituent communications (Fiorina 1977) the major conclusion is that, on average, a legislator is more likely to focus on non-ideological themes in communications such as claiming credit for new pork projects, or advertising that she helped a veteran get access to benefits than position taking. Fiorina sums up this argument,

“Congressmen know that the specific impact of broad national policies on their districts is difficult to see, that effects are hidden, so to speak. They know too that individual congressmen are not held responsible for the collective outcome produced by 535 members of Congress. Thus, in order to attain reelection, congressmen focus on things that are both more recognizable in their impact and more credible indicators of the individual congressman’s power - federal projects and individual favors for constituents. In order to purchase a steady flow of the latter, congressmen trade away less valuable currency - their views on public policy...

Overall then, programmatic activities are dangerous (controversial) on the one hand, and programmatic accomplishments are difficult to claim credit for, on the other. While less exciting, casework and pork barreling are both safe and profitable. For a reelection-oriented congressman the choice is obvious.”
If legislators are simply advertising or credit claiming in a majority of their messages, e-newsletters might not be ideologically informative. Since advertising is largely devoid of issue content – as everyone regardless of ideology should find these messages beneficial to their reelection goals – the texts should be similar irrespective of who sends them. While this is a prominent expectation in political communication research, this work was based on the Congress of the 1970s which is very different than the more polarized Congress of today (McCarty, Poole & Rosenthal 2006). The new polarized political environment may make different communication strategies more advantageous in ways that Mayhew (1974), Fiorina (1977), and Fenno (1978) could not foresee.

Another reason an underlying ideological continuum might not be apparent is if everyone is sending similar, anti-Congress messages. If everyone is “running against Congress” as posited by Fenno (1978) then communications may not be differentiating. This scenario would involve many negative e-mails disparaging Congress as an institution in an attempt to portray each individual legislator as better than the rest of his or her colleagues. Recent research has supported Fenno’s view that most legislators run against Congress (Bruff 1987, Patterson & Magleby 1992, Polsby & Schickler 2004). If all members are simply seeking to optimize their chances of being reelected, and if they think that advertising or running against Congress is a winning strategy, then their e-mails should be similar regardless of their ideology.

2.2 Ideological Communication

While most previous work indicates that legislators will not engage in a good deal of position taking in communications – for fear of alienating some voters – there are some reasons why
Legislators might use ideologically driven language in their e-newsletters.

Reputations are not built by actions such as votes alone; they can also made with the messages legislators send in e-mails, television interviews, C-SPAN coverage, and radio spots. Indeed, Newt Gingrich used C-SPAN at the start of his career to transform himself from a rookie backbencher to a rising conservative star by challenging Democratic leaders in a public manner from the floor go the House (Hagey 2011). Assuming each member exists in some ideological space in other public realms, there is no a priori reason that e-newsletters should be different and these messages provide an opportunity to legislators to develop an ideological reputation.

In developing a public persona, it is evident that legislators do not only talk about pork projects or constituent casework, they also take positions. E-newsletters should be consistent with other presentations of self made by a legislator and therefore there will likely be some level of position taking. Analyses of traditional franked mail, surveys, and interviews with legislative staff show that similar messages from newsletters are conveyed in personal appearances, media coverage, and other forms of communications with constituents (Lipinski 2004).

Legislators might send ideological e-mails because they perceive the audience of e-newsletters as more extreme than typical constituents. Legislators cannot send unsolicited e-newsletters, so everyone who receives e-newsletters must first sign up on the member’s website. Regarding campaign websites (rather than official office websites), there is evidence that engaged supporters visit candidates websites with greater frequency than other voters, despite the face that creators of campaign websites maintain that they are created for general and undecided voters (Druckman, Kifer & Parkin 2010). Campaign websites are different from official websites however in that all constituents may use the official websites to contact their members for help whether they support or oppose them once in office. To find out what types of people sign up for official e-newsletters, I placed a question on the 2012
Cooperative Congressional Election Study (CCES). I find that people who have subscribed to such communications are, on average, older, more educated, wealthier, more politically active, slightly more politically extreme, and are more likely to vote in primary elections, more likely to approve of their incumbent legislators, and think higher of Congress in general than non-subscribers.

Previous work using support vector machine analyses on House and Senate speeches shows that floor speeches are increasingly separable by ideology in more recent years (Yu & Diermeier 2010). Additionally, when the Wordscores technique is applied to House and Senate Speeches, the recovered orderings for individuals closely correlate with DW-NOMINATE scores (Beauchamp 2010). Yet, speeches are very different than e-newsletters. When members speak on the floor they are subject to party pressures, time limits, and germaneness rules. These speeches also occur in a directly political contact and by virtue of their setting and purpose, it is not surprising that researchers can recover an ideological continuum. What is yet to be seen is whether legislators continue the trend of politicized speech found inside the chambers of government in their constituent communications.

3 Communication Techniques

I have become convinced that scrutiny of purposive behavior offers the best route to an understanding of legislatures - or at least of the United States Congress. - Arnold in the Preface to Congress: The Electoral Connection, Mayhew 1974

Since the founding of the United States, with a variety of modifications over time, legislators have had the privilege of sending official mail to their constituents free of cost. In the past this was predominantly accomplished by use of the franking privilege applied to traditional “snail mail”. However, in recent times, legislators have also started to use online e-newsletters. Almost all legislators send e-mail communications to their constituents. In
the 111th Congress, 97 percent of House members and 85 percent of Senators had set up e-newsletters.  

E-newsletters are different than franked mail. Because they are opt-in subscriber based services, they are considered solicited messages and are exempted from pre-election restrictions (Glassman 2007). While they are still prepared by legislators in their official capacity and at taxpayer expense (as opposed to being prepared by a reelection campaign team) only 2 members of the 111th House include the disclaimer, “This mailing was prepared, published, and mailed at taxpayer expense.” Legislators include more pictures, partisan references, and biographical information in their official e-mails than allowed in their official traditional mail. If I find that legislator e-newsletters can be classified according to a recoverable underlying ideological dimension, this may signal that e-newsletters are a way to circumvent the restriction that taxpayer funded communications must be apolitical.

4 Data Description

All legislators are given resources to produce and maintain an official website. The data analyzed here are originally collected e-newsletters from August 2009 through November 2010 from all House and Senate members who offered such subscriptions on their official website. There were a total of 9,685 e-newsletters in this period. Legislator generated texts might more accurately reflect ideology or preferences of legislators than votes because they have 100% control of the content, whereas voting occurs under institutional constraints such as agenda control, party discipline measures and committee decisions.

Learning about legislator communications refines our comprehension of political behavior in a way that roll call voting cannot. Under normal electoral constraints, legislators need to cast votes even if they do not feel particularly compelled one way or another on an issue, or they risk being labelled a “shirker”. This compulsion is not the same for official
Legislators have complete control over the the content of the messages, and even whether to send them or not. There is no easily accessible registry where constituents can measure the verbosity of their legislator as compared to other legislators, like there is for voting records. Overall this data and the method of text analyses are unobtrusive and well suited to see if legislator communications reflect the ideological nature of voting.

4.1 Characteristics of the data

Legislator e-newsletters contain many types of messages. Table 1 shows the subject lines of a randomly generated sample of 10 e-newsletters, to demonstrate the variation in the messages.

[Table 1 about here.]

The e-mails vary in length, number topics covered, and focus, as is apparent from the above list. Some contain descriptions of governmental processes, position taking, credit claiming, advertising, clippings from local newspapers, local praise and many more topics.

As of November 2010, 97% of House members and 85% of Senators had some sort of e-newsletter sign up. 84% of House members and 77% of Senators send actual substantive e-newsletters beyond the initial “thank you for subscribing” form letters, I use these members for my analysis. Member ideology, as measured by the first and second dimension DW-NOMINATE is virtually uncorrelated with sending e-newsletters. The respective correlations are $\rho = -0.01$ and $\rho = -0.04$. Additionally, legislator extremity measured via squaring the DW-NOMINATE first dimension scores is also not related to whether a legislator sends e-newsletters ($\rho = 0.04$).
5 Methods: Wordscores

I use Laver, Benoit, and Garry’s (2003) Wordscores routine to recover positions of legislators by only using the information communicated to constituents in e-newsletters. Wordscores does not use information about word order, only the presence or absence and frequency of words are used to calculate point estimates and confidence intervals for each text (legislator). This approach is characterized as a “bag of words” and researchers in both computer science and political science have found that approaches like this are effective ways to analyze texts (Hopkins & King 2010).

The method requires information on pre-selected reference texts, and then uses that information to estimate positions on “virgin” texts. Wordscores is made up of two distinct procedures. After reading the text files and selecting reference texts and scores, (1) word “scores” are estimated for each word from individual word counts based on their relative frequency in the reference texts, (2) document scores are then estimated from a combination of component word scores for all virgin texts. The method relies on the probability of a document being a certain type depending on the presence or absence of words that are used in the reference texts. A simple, yet illustrative example of how Wordscores works is provided in the original Laver et al. 2003 article,

“As an example consider two reference texts, A and B. We observe that the word ‘choice’ is used 10 times per 10,000 words in Text A and 30 times per 10,000 words in Text B. If we know simply that we are reading the word ‘choice’ in one of the two reference texts, then there is a 0.25 probability that we are reading Text A and a 0.75 probability that we are reading Text B.”

Given this knowledge across all words, a word score is assigned for each word and then a combination is used to assign an overall text score. If Reference Text A is pinned at -1.0 and Reference Text B is pinned at 1.0 then the score for the example word ‘choice’ is given by the following:
0.25(-1.0) + 0.75(1.0) = -0.25 + 0.75 = +0.5.

After obtaining scores for every word in the reference text lexicon, virgin text scores are estimated by using the mean score for all the words within the text, weighted by their usage frequency.

Wordscores has previously been used to measure debates in parliaments (Meddaugh 2010) as well as political manifestos (Laver, Benoit & Garry 2003) and US floor speeches (Beauchamp 2010) with general success. Even with large and repeated violations of key methodological assumptions, Wordscores has repeatedly been shown to effectively sort parties and individuals in a variety of settings. One of the main critiques of the method is the measure of the returned intervals between individual members is not meaningful (Lowe 2008). In this paper I do not focus on the intervals per se, but rather the relationship of the overall ordering pattern with another established measure of ideology so this concern ought not present a challenge to my inferences.

5.1 Analysis

The guideline when choosing reference texts is that one must have access to confident estimates of, or assumptions about, positions on the policy dimension(s) under investigation (Laver, Benoit & Garry 2003). Using DW-NOMINATE scores based on roll calls in the 111th Congress, I select the 10 most liberal and 10 most conservative members of the 111th Senate and House that have substantive e-newsletters to use as reference texts. These rankings act as both a good estimate of where legislators are relative to each other and are less subjective than simply assuming who the most liberal and conservative legislators are. The members used for the reference texts are the “training” set for Wordscores and I report results for the remaining members of the Senate and House as those legislators are the “test”
set for the processing routine. The most liberal and most conservative members of the 111th Senate and House are presented in table 2.

I create one document for each legislator with the text of all of his or her e-newsletters sent from August 2009 - November 2010. The bolded members’ enewsletters in Table 2 are used as reference texts for this analysis. The concatenated collection of the most liberal members’ e-mails are scored -100 and the most conservative members’ e-mails are scored 100.

6 Results

The Wordscores estimates and DW-NOMINATE scores for the 111th Senate and House are displayed in figure 1; the Spearman’s rank order correlation coefficients appear in the lower left corner of each individual figure. The solid line through the scatter plots represents the fitted values of regressing the Wordscore estimates on the DW-NOMINATE. 95% confidence bands surround the line. DW-NOMINATE scores are used as a comparison because they are the most widely used measure of legislator ideology, and thus ideal to measure concordant validity of the resulting Wordscores estimates.

This analysis indicates that there is an underlying ideological dimension in the messages sent by legislators. The correlation between the Wordscore estimates and the first dimension DW-NOMINATE score is 0.69 and 0.73 for the Senate and House respectively. The correlation with the second dimension DW-NOMINATE scores are insignificant for both
chambers at -0.17 and 0.10 for the Senate and House respectively. The text measure captured in the Wordscores estimates better corresponds to the first dimension of DW-NOMINATE scores, widely interpreted as the economic dimension of politics, i.e. how much the government should be involved in the economy, than with the more nebulous second dimension. In the following sections I perform a series of variations on the Wordscores procedure to act as robustness checks on the main results reported above. These checks generally reinforce the finding that legislators communicate ideologically in their e-newsletters.

### 6.1 Within Party Results

These results are not simply generated by across party differences; looking within parties I still find a robust relationship between the text of e-newsletters and the first dimension DW-NOMINATE measure of ideology. This is important because it hedges against the possibility that the finding is based on a pattern of communication where each party issues talking points and specific terms to their legislators, and legislators then send them to constituents resulting in different Wordscores scores, by virtue of partisanship alone and not a more nuanced notion of ideology. From reading samples of the e-newsletter data, not all members of a party send the same messages, and having the results broken down by party makes the case for intra-party ideological ordering even stronger. Using the original reference texts, figures 2 and 3 show the within party correlation of the Wordscore estimates and the first and second dimension DW-NOMINATE scores for both Democrats and Republicans.

[Figure 2 about here.]

[Figure 3 about here.]

The first dimension DW-NOMINATE to Wordscore correlation for Democrats is 0.29 and 0.26 in the Senate and House respectively. For Republicans, the Wordscore to
first dimension DW-NOMINATE correlation is 0.31 and 0.60 for the Senate and House, but the coefficient for the Senate is non-significant. The story on the 2nd dimension DW-NOMINATE to Wordscore correlation remains the same in the Senate within the parties, no significant relationship is detected. In the House however, things are a bit different. As is apparent in the House graphs of figures 2 and 3, the second dimension is a more powerful separating force within parties than across parties. The resulting correlation for the Democrats is 0.32 and for the Republicans 0.29. The fact that there are strong Wordscore to second dimension ideology correlations is promising for future work on what exactly characterizes the second dimension. It has long been said to represent the social or regional aspects of politics, but an analysis of which words are most differentiating on the second dimension could help us better pin down what issues characterize this dimension.

6.2 Reciprocal Averaging

Wordscores is a supervised approach of text analysis because the researcher chooses the reference (training) texts using their knowledge of the dimension of interest. As a robustness test I applied an unsupervised iterative approach reciprocal averaging to the Senate e-newsletters. In this analysis each legislator is represented by a vector of the word frequencies in their e-newsletter texts just as with Wordscores, but unlike Wordscores no set of legislators are set aside as reference texts. The legislator vectors are stacked on top of each other so every row is a legislator, every column is a word, and the individual cells are the counts of word usage by each legislator. This results in an enormous matrix and reciprocal averaging attempts to summarize the data in fewer dimensions by finding the weighting that best describes the variation in relative frequencies of words across legislators. I estimate a one dimensional solution so that each legislator has a score that describes their position relative to others. The correlation between these resulting scores and DW-NOMINATE is 0.55. This lower than the Wordscore to DW-NOMINATE correlation, but is still a significant
relationship and shows that results reported above are not spuriously driven by the selection of the reference texts.

6.3 Random Reference Texts

As a falsification test I reran the entire analysis, but instead of using the 10 most liberal and conservative legislators for the reference texts I had the computer generate a random list of 20 legislator texts to use as the two reference texts. Choosing the references at random should result in a non-relationship with DW-NOMINATE scores. The results of this check are all non-significant as expected. In the Senate, the correlations with the first and second dimension DW-NOMINATE scores are 0.12 and -0.10 respectively, and the corresponding correlations in the House are 0.08 and 0.02. As is apparent from the resulting correlations, the Wordscores using the informed reference texts are not picking up an ideological ordering by some artifact.

Taken together, the data and analysis here indicate that the messages sent in official congressional e-newsletters are ideologically organized and they are similar to the existing measures of ideology that currently exist.\textsuperscript{18} These findings are at odds with the theoretical expectations put forward by dominant theories of political communications. Indeed, I find that legislators are not just sending advertising or credit claiming messages devoid of political content, but rather that there is an ideological underpinning to the types of topics they discuss and the words used to describe those topics. In the next section, I explore in more detail the differences between liberals and conservatives in how they engage constituents in e-newsletters.
7 Ideological Content

Table 3 displays the most frequent words that liberal and conservative legislators use in e-newsletters. General e-mail specific words have been removed such as “subject”; “recipient”, “sender” because these are not choice words used by legislators, but rather necessary given by e-mail clients. This table indicates, unsurprisingly, that Senators are prone to refer to themselves and others in their institution as “Senate” and “Senator” are both high ranking words. Both liberal and conservative legislators write about health or health care more often than other things during the 111th Congress in which the Patient Protection and Affordable Care Act was passed. Beyond self-referencing and health care, the other frequent words favor funding and grants for liberal legislators and discussion of Washington and bills for conservatives.

Table 3 shows the most frequent words, but another way to understand legislator communication is to look at words they use together more often than they use them apart. These words make up phrases or terms of art, also known as collocations. I have created a list of the top 200 collocations for the Senate and House reference texts to get a better sense for the content liberals and conservatives emphasize in their messages. There are some collocations that are shared by both liberals and conservatives, in the Senate the most frequent of these terms are: “health care”, “the senate”, “washington dc”, “department of”, “united states”. In the House the common collocations are “health care”, “U. S.”, “will be”, “the house”, “D. C.”. In order to get a glimpse of what liberals talk about versus conservatives, the unique collocations are of interest. Table 4 shows some the most popular unique liberal and conservative collocations of the Senate and House.
All of the topics indicated by the most frequent common and unique collocations were timely during the 111th Congress. Liberals from the Senate stressed Catalog of Federal Domestic Assistance (CFDA) grants and programs - mostly administered through the Department of “Health and Human Services” that their constituents may be able to take advantage of. Senate conservatives focused more heavily on health care reform being pushed by “the President’ over the will of the “American people”. In the House, liberal communication is more varied. Liberal House legislators focused on “social security” as a frame for greater government involvement in health care while railing against practices of “insurance companies”. Conservatives in the House, like those in the Senate wrote about “the government” run take over of health care saying “they are” (the administration) ignoring the desires of the “American people”.

The most popular collocations are no doubt a rough way to summarize message content, but they provide a quick and broad understanding of what legislators are writing about. While there is overlap in the use of popular collocations by conservatives and liberals, there are also notable differences that indicate how legislators focus on different issues and use different language when communicating to their constituents.

8 Implications and Conclusions

Many scholars conclude that polarization in various forms has been on the rise since the 1970s. Among ordinary individuals (Levendusky 2009, Hetherington & Weiler 2009); partisans (Abramowitz & Saunders 2005, Evans 2003, Layman, Carsey & Horowitz 2006); elites (Abramowitz & Saunders 2008); and elected politicians (McCarty, Poole & Rosenthal 2006). Even those who believe the electorate has not become more polarized over time concede that politicians and strong partisans of the major parties have (Fiorina, Abrams & Pope 2006). Furthermore, divisions in the political sphere that result from polarization are problem-
atic because they may usher in underrepresentation or exclusion of certain groups from the political process (Baldassarri & Gelman 2008).

In this paper, I present data that speaks to current levels of ideological language in the way legislators communicate to their constituents. In order to focus on representative purposive behavior, I use official messages. I find that there is a level of message differentiation across legislators that follows the ideological differentiation we observe in voting patterns. This is an important addition to previous findings regarding the polarization of voting behavior. Although this dataset cannot yet speak to over time trends, it fits as another piece in the puzzle. Liberal legislators communicate to their constituents in a manner different than conservative legislators do and these trends mirror the differences observed in voting behavior. As time passes, the data used here will be suited to answer more longitudinal questions regarding increasing or decreasing patterns of ideologically tractable language in legislator communications.21

This paper analyzes the communication behaviors of legislators in the act of representation.22 A better understanding of what legislators do once in office is critical to cultivating a better understanding of representation and is necessary to test theories of political communication. These results indicate that House members speak to their constituents in ideological ways, not in the policy-devoid language of name-based advertising as expected by some of the most widely posited theories of Congressional communications. Additionally, this paper establishes an empirical basis for political theorists to interrogate the question of whether these results are good or bad for democratic responsiveness and electoral fairness. Fenno state that, “A complete treatment of explanatory styles would monitor the totality of congressional communications practices.” This paper certainly does not meet that threshold, but it is an importation addition. With work on other forms of communication, press releases (Grimmer & King 2010, Grimmer forthcoming), traditional mail (Hickey 2010), twitter (Williams & Gulati 2010, Sparks 2010) the discipline is collectively enhancing our
understanding of the political process more and more each day. A good deal of understanding of how elected officials interact with voters comes from examining a small number of legislators via Fenno’s work in Home Style (1978). By advancing that line of research with a new data source representing the universe of all e-newsletters from a given time period, I hope to draw a more complete picture of the representative process in the United States.
Table 1: Random Selection of 10 E-Newsletter Senders and Subject Lines (111th Congress)

<table>
<thead>
<tr>
<th>Sender</th>
<th>Subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>Senator Richard Burr</td>
<td>A Message from Senator Richard Burr</td>
</tr>
<tr>
<td>Congressman Dave Reichert</td>
<td>Promoting solutions for health care</td>
</tr>
<tr>
<td>Congressman Doc Hastings</td>
<td>Doc Hastings HANFORD UPDATE E-Newsletter</td>
</tr>
<tr>
<td>Congresswoman Ann Kirkpatrick</td>
<td>Join Me for a Tele-Town Hall on the Second Amendment</td>
</tr>
<tr>
<td>Congressman Alm Joseph Cao</td>
<td>Update: Cao Calls on House Colleagues to Consider Raising Oil Company Liability Cap for Spills</td>
</tr>
<tr>
<td>Senator Robert Menendez</td>
<td>Senator Menendez’s E-Newsletter: The Menendez Message</td>
</tr>
<tr>
<td>Congressman Tim Murphy</td>
<td>E-News from Congressman Murphy</td>
</tr>
<tr>
<td>Senator Lamar Alexander</td>
<td>eNewsletter Volume 7, Issue 8</td>
</tr>
<tr>
<td>Senator Barbara Boxer</td>
<td>Senator Boxer Visits the Center for Employment Training</td>
</tr>
<tr>
<td>Congressman Mark Kirk</td>
<td>Happy Thanksgiving and Farewell to the House</td>
</tr>
</tbody>
</table>
Table 2: Most Liberal and Most Conservative Members of the Senate and House (111th Congress, DW-NOMINATE Scores)

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Most Liberal</th>
<th>State</th>
<th>Most Conservative</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Goodwin</td>
<td>WV</td>
<td>Coburn</td>
<td>OK</td>
</tr>
<tr>
<td>2</td>
<td>Sanders</td>
<td>VT</td>
<td>Demint</td>
<td>SC</td>
</tr>
<tr>
<td>3</td>
<td>Whitehouse</td>
<td>RI</td>
<td>Inhofe</td>
<td>OK</td>
</tr>
<tr>
<td>4</td>
<td>Harkin</td>
<td>IA</td>
<td>Bunning</td>
<td>KY</td>
</tr>
<tr>
<td>5</td>
<td>Kirk</td>
<td>MA</td>
<td>Ensign</td>
<td>NV</td>
</tr>
<tr>
<td>6</td>
<td>Durbin</td>
<td>IL</td>
<td>Kyl</td>
<td>AZ</td>
</tr>
<tr>
<td>7</td>
<td>Burris</td>
<td>IL</td>
<td>Enzi</td>
<td>WY</td>
</tr>
<tr>
<td>8</td>
<td>Brown</td>
<td>OH</td>
<td>Barrasso</td>
<td>WY</td>
</tr>
<tr>
<td>9</td>
<td>Lautenberg</td>
<td>NJ</td>
<td>Vitter</td>
<td>LA</td>
</tr>
<tr>
<td>10</td>
<td>Boxer</td>
<td>CA</td>
<td>Sessions</td>
<td>AL</td>
</tr>
<tr>
<td>11</td>
<td>Gillibrand</td>
<td>NY</td>
<td>Burr</td>
<td>NC</td>
</tr>
<tr>
<td>12</td>
<td>Franken</td>
<td>MN</td>
<td>Cornyn</td>
<td>TX</td>
</tr>
<tr>
<td>13</td>
<td>Reed</td>
<td>RI</td>
<td>Risch</td>
<td>ID</td>
</tr>
<tr>
<td>14</td>
<td>Levin</td>
<td>MI</td>
<td>Crapo</td>
<td>ID</td>
</tr>
<tr>
<td>15</td>
<td>Akaka</td>
<td>HI</td>
<td>Chambliss</td>
<td>GA</td>
</tr>
<tr>
<td>16</td>
<td>Leahy</td>
<td>VT</td>
<td>McConnell</td>
<td>KY</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Most Liberal</th>
<th>State</th>
<th>Most Conservative</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>McDermott</td>
<td>WA</td>
<td>Paul</td>
<td>TX</td>
</tr>
<tr>
<td>2</td>
<td>Kucinich</td>
<td>OH</td>
<td>Sensenbrenner</td>
<td>WI</td>
</tr>
<tr>
<td>3</td>
<td>Stark</td>
<td>CA</td>
<td>Flake</td>
<td>AZ</td>
</tr>
<tr>
<td>4</td>
<td>Woolsey</td>
<td>CA</td>
<td>Stutzman</td>
<td>IN</td>
</tr>
<tr>
<td>5</td>
<td>Lee</td>
<td>CA</td>
<td>Broun</td>
<td>GA</td>
</tr>
<tr>
<td>6</td>
<td>Miller</td>
<td>CA</td>
<td>Graves</td>
<td>GA</td>
</tr>
<tr>
<td>7</td>
<td>Filner</td>
<td>CA</td>
<td>Chaffetz</td>
<td>UT</td>
</tr>
<tr>
<td>8</td>
<td>Waters</td>
<td>CA</td>
<td>Shadegg</td>
<td>AZ</td>
</tr>
<tr>
<td>9</td>
<td>Olver</td>
<td>MA</td>
<td>Mcclintock</td>
<td>CA</td>
</tr>
<tr>
<td>10</td>
<td>Conyers</td>
<td>MI</td>
<td>Campbell</td>
<td>CA</td>
</tr>
<tr>
<td>11</td>
<td>Hinchey</td>
<td>NY</td>
<td>Hensarling</td>
<td>TX</td>
</tr>
<tr>
<td>12</td>
<td>Lewis</td>
<td>GA</td>
<td>Lummis</td>
<td>WY</td>
</tr>
<tr>
<td>13</td>
<td>Grijalva</td>
<td>AZ</td>
<td>Franks</td>
<td>AZ</td>
</tr>
<tr>
<td>14</td>
<td>Frank</td>
<td>MA</td>
<td>Rohrabacher</td>
<td>CA</td>
</tr>
<tr>
<td>15</td>
<td>Ellison</td>
<td>MN</td>
<td>Pence</td>
<td>IN</td>
</tr>
</tbody>
</table>

**Bolded** indicates substantive e-newsletters used as reference texts

Data from Voteview.com.
Table 3: Most Used Words of the Senate and House E-newsletters (111th Congress)

<table>
<thead>
<tr>
<th>Word</th>
<th>Liberal % of all Words</th>
<th>Conservative Word</th>
<th>Conservative % of all Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Senator</td>
<td>0.008</td>
<td>Senate</td>
<td>0.007</td>
</tr>
<tr>
<td>Program</td>
<td>0.006</td>
<td>Senator</td>
<td>0.004</td>
</tr>
<tr>
<td>Health</td>
<td>0.005</td>
<td>Health</td>
<td>0.004</td>
</tr>
<tr>
<td>Senate</td>
<td>0.005</td>
<td>Bill</td>
<td>0.004</td>
</tr>
<tr>
<td>Funding</td>
<td>0.004</td>
<td>Washington</td>
<td>0.004</td>
</tr>
<tr>
<td>Grants</td>
<td>0.004</td>
<td>Care</td>
<td>0.004</td>
</tr>
<tr>
<td>Information</td>
<td>0.004</td>
<td>Please</td>
<td>0.003</td>
</tr>
<tr>
<td>Office</td>
<td>0.004</td>
<td>DC</td>
<td>0.003</td>
</tr>
<tr>
<td>Please</td>
<td>0.004</td>
<td>Office</td>
<td>0.003</td>
</tr>
<tr>
<td>Current</td>
<td>0.003</td>
<td>Reform</td>
<td>0.003</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Word</th>
<th>Liberal % of all Words</th>
<th>Conservative Word</th>
<th>Conservative % of all Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health</td>
<td>0.005</td>
<td>House</td>
<td>0.006</td>
</tr>
<tr>
<td>Congressman</td>
<td>0.004</td>
<td>Please</td>
<td>0.005</td>
</tr>
<tr>
<td>Care</td>
<td>0.003</td>
<td>Congressman</td>
<td>0.004</td>
</tr>
<tr>
<td>House</td>
<td>0.003</td>
<td>Health</td>
<td>0.004</td>
</tr>
<tr>
<td>Act</td>
<td>0.003</td>
<td>Bill</td>
<td>0.004</td>
</tr>
<tr>
<td>Bill</td>
<td>0.003</td>
<td>Care</td>
<td>0.004</td>
</tr>
<tr>
<td>Jobs</td>
<td>0.003</td>
<td>Office</td>
<td>0.004</td>
</tr>
<tr>
<td>Veterans</td>
<td>0.003</td>
<td>Government</td>
<td>0.003</td>
</tr>
<tr>
<td>New</td>
<td>0.003</td>
<td>Congress</td>
<td>0.003</td>
</tr>
<tr>
<td>Insurance</td>
<td>0.003</td>
<td>Washington</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Stop words and general e-mail words excluded.
Table 4: Most Used Uniquely Liberal and Conservative Collocations of the Senate and House E-newsletters (111th Congress)

<table>
<thead>
<tr>
<th>Liberal</th>
<th>Conservative</th>
<th>Liberal</th>
<th>Conservative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Senate</td>
<td>House</td>
<td>Senate</td>
<td>House</td>
</tr>
<tr>
<td>CFDA number</td>
<td>Care Reform</td>
<td>Social Security</td>
<td>Member Of</td>
</tr>
<tr>
<td>States Senator</td>
<td>The Bill</td>
<td>Care Reform</td>
<td>The Government</td>
</tr>
<tr>
<td>Health And</td>
<td>This Week</td>
<td>Department Of</td>
<td>There Is</td>
</tr>
<tr>
<td>And Human</td>
<td>The United</td>
<td>Insurance Companies</td>
<td>They Are</td>
</tr>
<tr>
<td>Human Services</td>
<td>The President</td>
<td>Last Month</td>
<td>There Are</td>
</tr>
<tr>
<td>US Senate</td>
<td>The State</td>
<td>Continue To</td>
<td>American People</td>
</tr>
<tr>
<td>Dear Friend</td>
<td>American People</td>
<td>Wall Street</td>
<td>That Is</td>
</tr>
<tr>
<td>My Work</td>
<td>An Amendment</td>
<td>Access To</td>
<td>Care Bill</td>
</tr>
<tr>
<td>Grants For</td>
<td>I Will</td>
<td>Able To</td>
<td>Federal Government</td>
</tr>
<tr>
<td>The Senator</td>
<td>US Sen</td>
<td>Funding For</td>
<td>All Of</td>
</tr>
</tbody>
</table>

Stop words and general e-mail collocations excluded.
Figure 1: 111th Congress

Senate Wordscores v. 1st Dimension DW-NOMINATE

Senate Wordscores v. 2nd Dimension DW-NOMINATE

House Wordscores v. 1st Dimension DW-NOMINATE

House Wordscores v. 2nd Dimension DW-NOMINATE

** p < 0.05  * p < 0.10
Figure 2: 111th Congress (Democrats Only)

Senate Wordscores v. 1st Dimension DW-NOMINATE

Senate Wordscores v. 2nd Dimension DW-NOMINATE

House Wordscores v. 1st Dimension DW-NOMINATE

House Wordscores v. 2nd Dimension DW-NOMINATE

\( \rho: 0.29^* \ N: 37 \)

\( \rho: 0.18 \ N: 37 \)

\( \rho: 0.26^{**} \ N: 190 \)

\( \rho: 0.32^{**} \ N: 190 \)

\** p < 0.05 \ * p < 0.10
Figure 3: 111th Congress (Republicans Only)
A Appendix

10 legislators are selected to make the reference texts. The following figures shows how four measures vary over a different amount of legislators included to construct the reference texts. The references were determined by starting with the most extreme legislator on each end of Optimal Classification rankings and then adding the next closest legislator.

Measure Changes Across Varying Numbers of legislators in Reference Texts

On the x-axis is the number of reference texts. The upper left graph shows that the precision of the resulting Wordscore estimate is increasing (the standard error decreases) as the number of references increases. The upper right graph displays a measure of comprehensiveness of the reference text, namely the average amount of words that can be scored using a given number of legislators in the reference texts. Any words that do not occur in the reference texts cannot be used to scale the virgin texts, so a higher number is better.
As the figure shows this number is unsurprisingly increasing with the number of legislators used in the reference texts. The lower portion of the figure relates the substantive findings of this paper to the number of legislators used in the reference texts. Generally speaking, more legislators in references increases the correlation with 1st dimension NOMINATE scores, but the picture is less clear cut regarding the 2nd Dimension, indicating the need for a further analysis.

B Appendix

In order to test for a more complicated relationship I estimate three simple models, one linear and two polynominal models of the following form:

\[
\text{Wordscore}_i = \alpha + \beta(NOM_i) + \epsilon_i \\
\text{Wordscore}_i = \alpha + \beta(NOM_i) + \beta_2((NOM_i)^2) + \epsilon_i \\
\text{Wordscore}_i = \alpha + \beta(NOM_i) + \beta_2((NOM_i)^2) + \beta_3((NOM_i)^3) + \epsilon_i
\]

Where i indexes each legislator and NOM is the first dimension 111th Congress DW-NOMINATE score for each i legislator. Figure 5 displays the results from models 1-3. In the upper left graph of the linear model, a loess curve is overlaid in red. These results suggest that the relationship between the estimated Wordscores and first dimension DW-NOMINATE are best characterized as linear or cubic. Comparing the models indicates that the R-squared does not increase by much when going from a linear to quadratic (0.001) but going to a cubic model does offer a bit of a better fit with an R-squared improvement of 0.041.
In the name of parsimony, and noting the appearance of the locally weighted (lowess) regression curve, one could argue for a linear interpretation of the relationship between legislator Wordscores and DW-NOMINATE score. On the other hand, an F test does signal that the cubic model does significantly add to the explanation of the variance, and is a better model than a linear model for these data. While the linear model might not be the best technical fit of the data it is still the most straightforward way to think about the relationship between the two variables of interest, and therefore I am comfortable with this assumption underlying the use of Wordscores.

Notes


2 The texts analyzed here are direct legislator-to-constituent communications rather than press releases. These media differ in target audience, method of delivery, and subject matter.


4 The CCES draws from a nationally representative survey population for the full survey. The specific question posed to a randomly selected 1,000 respondents was, “Have you ever subscribed to e-mail updates such as an e-newsletter or Real Simple Syndication (RSS) feed from any of the following elected officials?” followed by the name of their Representative, junior Senator, and senior Senator.

5 Support Vector Machine analysis is similar to the Wordscores method I apply in this paper. Briefly, both methods are probabilistic supervised learning methods that rely on training/reference data (words within texts) and user supplied classifications to then map test/virgin data to the classification scheme based on which words are most often associated with the original classification scheme for the training texts. The differences are largely in computational implementation, but results using either of these methods on the same data produces similar results (Hillard, Purpura & Wilkerson 2007, Beauchamp 2010).

6 This amount of functioning newsletters greatly exceeds that observed in state legislative bodies (Herrnson, Stokes-Brown & Hindman 2007).

7 The opt-in characterization of legislator e-mail can be somewhat disputed. Citizens can be subscribed in the normal way by going to a legislator website and explicitly signing up to receive e-mail updates, but other more covert sign up methods also exist. If a citizen e-mails an legislator on some issue, or takes a survey on the legislator’s website in which they provide their e-mail address, those addresses are sometime captured and then added to the list of citizens who explicitly signed up to receive e-newsletter updates. Additionally, some offices keep e-mail addresses from term to term even if a different legislator has been elected to a seat. Some members of the House that later get elected to the Senate also keep the e-mail addresses of citizens who subscribed to their House e-newsletters and then use those addresses when sending their Senate e-newsletters.

8 This disclaimer is required on all traditional mail sent with the franking privilege. No senators include
this in their official e-mails.

9 Official websites are those maintained and linked from www.senate.gov and www.house.gov. These are different from campaign websites which are maintained by individual members’ campaigns.

10 Substantive messages are sent by 81% of House Democrats and 88% of House Republicans do so. This difference is not statistically different from zero.

11 Discarding word order is standard in most textual analyses. Differentiation based on the rate of occurrence of words is typically enough to signal divergences in the latent meanings of texts.

12 DW-NOMINATE scores obtained from voteview.com. Appendix A shows the performance statistics using various numbers of legislators for the reference texts.

13 This compilation excludes the typical form “Thank you for signing up” letters. Each text is stripped of punctuation and all words are made lowercase. These are the simple preprocessing steps employed for standard text analysis procedures (Grimmer & King 2010).

14 Note that, DW-NOMINATE scores are pinned such that -1 is liberal and 1 is conservative.

15 One House member, Ruppersberger (MD-2) is omitted from the graphs in order to better render them. His Wordscore estimate is 49.63 with a standard error of 0.15 and was used in calculating the correlations and regression lines.

16 Of course the relationship between the Wordscores and first dimension DW-NOMINATE score need not be linear. It could be the case, that as members get more ideologically extreme, their rate of using ideological language increases at a steeper rate and a quadratic model or even a higher ordered model would better describe this relationship I address this in appendix B.

17 Spearman’s rho throughout.

18 An additional benefit from the Wordscores technique as applied to e-newsletters is the ability to place members that do not or have not yet placed votes in the House to have DW-NOMINATE scores. In this sample Pedro Pierluisi, the representative from Puerto Rico, Gregorio Kilili Camacho Sablan, the representative from the Northern Mariana Islands, Donna Christensen, the representative from the US Virgin Islands and Ted Deutch, the new FL-19 representative do not have 111th Congress DW-NOMINATE scores, yet they send e-newsletters so they can be placed using Wordscores. Christensen, Kilili Camacho Sablan, and
Pieluisi, are slightly left of center (the “center” or average of all reported wordscores for the House is -1.21.) with wordscores and (standard errors) of -5.01 (1.00); -5.29 (0.20); and -13.50 (0.75) respectively. On the more conservative side is the newly elected, Deutch with a wordscore of 2.60 and (0.70) for his standard error.

19 “Stop words” are excluded. Stop words are words such as ‘a’, ‘the’, ‘and’ etc, very common words that are necessary to construct sentences in english, but ones that do not necessary convey information beyond their grammatical utility. Stop word list provided by Onix at http://www.lextek.com/manuals/onix/stopwords1.html. E-mail list author generated.

20 Many collocations in natural language texts are simple words that commonly appear together like “of the” or “in the”; these words are not presented here as they do nothing to give context to the documents. I have also eliminated common e-mail and sender collocations such as “click here” or “mail updates” or “barbara boxer” for the same reason.

21 In the future I intend to classify and analyze different message types, i.e. position taking, advertising, and credit claiming separately.

22 As opposed to the papers that describe legislator communication in the act of campaigning (Herrnson, Stokes-Brown & Hindman 2007, Druckman, Kifer & Parkin 2010).

References


Beauchamp, Nick. 2010. “Text-Based Scaling of Legislatures: A Comparison of Methods with Applications to the US Senate and UK House of Commons.”


36


