

**Dangers, Toils, and Snares:**  
U.S. Senators' Rhetoric of Public Insecurity and Religiosity

Emily Gade<sup>1</sup>  
Sarah Dreier<sup>2</sup>  
Jon Schaeffer<sup>3</sup>  
John Wilkerson<sup>4</sup>  
Anne Washington<sup>5</sup>

---

<sup>1</sup>Corresponding Author: [ekgade@uw.edu](mailto:ekgade@uw.edu); *University of Washington*

<sup>2</sup>*University of Washington*

<sup>3</sup>*University of Washington*

<sup>4</sup>*University of Washington*

<sup>5</sup>*NYU Steinhardt*

### **Abstract**

Nearly all members of the U.S. Congress claim a religious affiliation. Personal religious beliefs shape politicians' candidacy, policy priorities, and congressional voting behavior. Yet we know relatively little about elected officials' public-facing religiosity. How, and in what contexts, do congressmembers publicly invoke religious rhetoric in their constituent communications? Congressmembers could use religious rhetoric alongside optimistic dispositions of stability and prosperity (H1), or to quell feelings of insecurity, threat, and anxiety (H2). To test these hypotheses, we use the Internet Archive's 90-terabyte collection of content from the U.S. government's Internet domain. We retrieve and analyze this archive's web-crawl captured text from U.S. senators' official congressional websites during four election years (2006, 2008, 2010, and 2012). We use dictionary methods to construct a measure of a senator's public-facing religiosity: her proportional use of religious terms in a given year. We then use a beta regression to estimate the relationship between a senator's use of optimistic and anxiety-related terms in a given year (each as a proportion of her overall yearly words) and her public-facing religiosity. We find that senators who publicly display anxious dispositions in a given year, regardless of real anxiety-inducing events, are strongly likely to invoke public-facing religiosity. Substantively, this research introduces a generalizable, time-variant, publicly relevant measure of congressmembers' religiosity and suggests crucial links between expression of religious ideology and insecurity-based social anxieties. We also make an invaluable applied methodological contribution: By employing web-based archival collections, big-data management, text-data dictionary analysis, and statistical analysis, we model the promise truly "big" data sources offer for understanding political dynamics.<sup>6</sup>

---

<sup>6</sup>This research is an outgrowth of work conducted under NSF #1243917 and builds on work conducted in the eScience Big Data Incubator at the University of Washington (2014).

*Through many dangers, toils and snares we have already come. T'was grace that brought us safe thus far, and grace will lead us home.* (“Amazing Grace” Christian Hymn, 1779)

## 1 Introduction

Despite constitutional prohibitions against religious requirements for elected public office, virtually all current and former members of the U.S. Congress self-identify as religious. Personal religious beliefs shape politicians’ candidacy, policy priorities, and congressional voting behavior. Similarly significant is their *public* mobilization of religion. Yet researchers know relatively little about this **public-facing religiosity**: elected officials’ invocation of religious concepts in their official communications with constituents. How, and in what contexts, do congressmembers publicly invoke religious rhetoric in their constituent communication and engagement?

Religious rhetoric provides a bridge between material policy circumstances and the symbolic ideas and emotions that confer meaning upon—and suggest ramifications of—those circumstances (Geertz, 2009). Since Abrahamic monotheism predominates American religiosity, these symbolic ideas most often take the form of a benevolent God who facilitates peace and prosperity and who provides comfort and security amid turmoil. Accordingly, members of the U.S. Congress may use religious concepts in times of perceived stability to invoke optimistic sentiments of gratitude, hope, or aspiring prosperity. Alternatively, they may use religion to quell feelings of insecurity, threat, and anxiety.

This article seeks to develop a measure of U.S. senators’ public-facing religiosity and to identify which dispositional attitudes (optimism or anxiety) religious rhetoric most frequently accompanies. To do so, we examine the content of U.S. senators’ official congressional websites, according to web-crawl captures during four election years (2006, 2008, 2010, and 2012). In each of these years, the Library of Congress contracted with the Internet Archive to target congressional websites as seed uniform resource locators (URLs). This election-year targeting strategy dramatically increased the quality of this data. We scraped this 90-terabyte collection of archived content for terms that the Linguistic Inquiry and Word Count (LIWC) dictionary categorizes as optimistic, anxiety, and religious terms.<sup>7</sup>

Using this big-data source, we construct a measure of senators’ proportional use of religious terms. We ground this measure of public-facing religiosity with analyses from the U.S. House of Representatives (2012). This verifies the external validity of our web-crawled data, word lists, and proportional-use measures. Next, we use a beta regression to estimate the relationships between a senator’s proportional use of optimistic and anxiety terms in a given year (explanatory variables) and her proportional use of religious terms that year (outcome variable).

This analysis yields three main results. First, conservatives, men, and senators representing high-religiosity constituencies adopt more prominent public-facing religiosity than their liberal, female counterparts who represent less-religious voters. Second, while optimism and anxiety both predict the frequency of religious terms, the substantive effect of anxiety on religious rhetoric is double the association between optimism and religion. As a senator adopts increasingly anxious rhetorical dispositions, she becomes substantially more likely to invoke religious rhetoric. Third, although anxiety terms covary with religious rhetoric, they are *not*

---

<sup>7</sup>We also use the U.S. Department of Homeland Security’s security-related “watch” word lists.

associated with actual contexts of insecurity (e.g., terrorist attacks at the state or natural disasters in the state). Real events do, however, have a negative relationship to senators' use of optimistic rhetoric. This suggests that senators promote unfounded dispositions of insecurity while offering religiously intertwined policy solutions to alleviate constituents' concerns during times of perceived insecurity.

This research makes several substantive and applied methodological contributions to social science research. We create a groundbreaking measure of public-facing religiosity among U.S. senators and representatives. This measure arguably has greater social significance than quantifications of elected officials' *personal* religious activity alone (see, for example, Guth (2014)'s valuable study of U.S. representatives). By identifying a positive association between perceptions of insecurity and public-facing religious rhetoric, we contribute substantive insights about religion's functional role in American political communication. Methodologically, this research exploits a truly "big"—and largely unused—data source to categorize, quantify, and analyze senators' public communication strategies and to extrapolate their approaches to public policy. It employs a unique combination of web-based archival collection, big-data management, text and statistical analysis. In doing so, it models a rigorous approach to leveraging the tremendous breadth of Internet data in order to advance social science research in the digital age. Indeed, these innovative approaches to catalyzing big, messy data is in the vanguard of political science.

The remainder of this article proceeds as follows. Section Two discusses religion's relevance in the U.S. Congress. Section Three hypothesizes connections between public optimism, anxiety, and religious rhetoric. Section Four discusses the Internet Archive's ".gov" data source, our scraping of this data for pre-established word lists, our construction of novel proportional-rhetoric variables, and the strengths and limitations of web-based archive data for social science analysis. Section Five presents regression model results and several robustness checks that substantiate our conclusions. Section Six closes with a discussion of this research's substantive and applied methodological contributions to political science and discusses avenues for future research.

## 2 Religion's Relevance in the U.S. Congress

Article 6 of the U.S. Constitution directs that "no religious test shall ever be required as a qualification to any office or public trust under the United States." This federal prohibition<sup>8</sup> of a religious metric for elected office has not prevented religious belief—whether authentic or feigned (Alberts, 2008)—from often serving as a *de facto* prerequisite for candidates seeking public office. Although twenty percent of American adults today claim no religious affiliation, virtually all members of the U.S. Congress self-identify as religious. To date, Congress has contained only one open atheist.<sup>9</sup> The current 115th U.S. Congress has one explicitly religiously unaffiliated member<sup>10</sup> and ten additional members (all Democrats) who do not report a religious

<sup>8</sup>Seven U.S. states explicitly prohibit those who deny the existence of a God from holding state office: Arkansas, Maryland, Mississippi, North Carolina, South Carolina, Tennessee, and Texas. Pennsylvania stipulates that no believer's religious sentiments may disqualify her from holding office but does not extend this protection to atheists (Schwarz, 2014).

<sup>9</sup>Rep. Pete Stark (D-CA), who served from 1973-2013, announced his atheism in 2007. He was defeated by another Democrat in California's 2012 general election.

<sup>10</sup>Rep. Kyrsten Sinema (D-AZ) avoids labeling herself a non-believer, clarifying instead that she "does not consider herself a member of any faith community" (Wing, 2017).

affiliation (Sandstrom, 2017).<sup>11</sup> Nor does Article 6 prevent religious beliefs from motivating U.S. voters' decisions or their elected officials' legislative behavior. In fact, many of the U.S. founders believed that religious (specifically Christian) morality was essential to a successful democracy (Waldman, 2009, 59-61).<sup>12</sup> They frequently recalled Biblical references in their political rhetoric (Dreisbach, 2011).

Today, religious candidates, especially Christians, are elected at disproportionately high rates.<sup>13</sup> American voters may gravitate toward religious politicians because the U.S. population remains highly religious, even as religiosity has dwindled in western Europe and other industrialized nations (Jenkins, 2011, 113), (Norris and Inglehart, 2011, 84-5).<sup>14</sup> Among the highly religious U.S. citizenry, voters' religious beliefs routinely shape their public opinions and their voting behavior (Smidt, Guth and Kellstedt, 2017).<sup>15</sup> In general, voters perceive religious candidates to be more trustworthy than their non-religious counterparts (Clifford and Gaskins, 2016, 1066). By professing a shared religious affiliation or a comparable belief in God, a political candidate may signal to her voters and constituents that she will advance their policy preferences or ideological priorities in Congress.

Once elected, legislators' own religious affiliation (Fastnow, Grant and Rudolph, 1999; Page et al., 1984), their degree of orthodoxy (Green and Guth, 1991), the salience of own their religious beliefs (Yamane and Oldmixon, 2006), and the religious breakdown of their constituency (Marchetti and O'Connell, 2017; Meier and McFarlane, 1993; O'Connor and Berkman, 1995; Oldmixon, 2005)<sup>16</sup> all shape legislators' voting behavior and policy preferences. This is particularly true among hot-button private issues (Oldmixon and Hudson, 2008), including abortion (Chressanthi, Gilbert and Grimes, 1991; Gohmann and Ohsfeldt, 1994; Richardson and Fox, 1972, 1975; Schechter, 2001; Tatalovitch and Schier, 1993) and marriage equality (Blackstone and Oldmixon, 2015; Guth, 2014).<sup>17</sup> However, with the polarization of American politics (Layman, 1999) and Christian conservatism's entrenchment in the Republican Party,<sup>18</sup> these effects are increasingly mediated through congressmembers' partisan identifications<sup>19</sup> and po-

<sup>11</sup>Senators Baldwin (D-WI), Bennet (D-CO) and Duckworth (D-IL) and Representatives Blumenauer (D-OR), Bonamici (D-OR), Chu (D-CA), Foster (D-IL), Huffman (D-CA), Jayapal (D-WA), and Pocan (D-WI) provide no religious affiliation.

<sup>12</sup>Four decades after the U.S. Constitution took effect, Tocqueville wrote that Judeo-Christian morality provided a crucial counter to the unwieldy, individualist-oriented approach to democracy he observed in the United States. He explicitly excluded Islam as incompatible with democracy (de Tocqueville, 1835).

<sup>13</sup>While 71 percent of the American public identify as Christian, 90.7 percent of Congress is Christian. This number has remained largely stable since 1979. Judaism is slightly overrepresented (by 3.6 percent) in Congress (Sandstrom, 2017).

<sup>14</sup>Some scholars suggest that religiosity has continued to thrive in the United States precisely because its absence of a state-sponsored religion enabled a competitive religious marketplace that helped maintain religion's public relevancy (Stark and Iannaccone, 1994; Stark and Finke, 2000).

<sup>15</sup>Thirty-eight percent of Americans surveyed in 2003 reported that their religious beliefs have at least an occasional influence on their voting decisions (Pew Forum on Religion & Public Life, 2003).

<sup>16</sup>Many seek to represent their constituents' perspectives and ideologies. For example, members of Congress appear positively responsive to their responsibility to represent Muslim constituents (Martin, 2009).

<sup>17</sup>Newman et al. (2016) identifies several other policy areas influenced by policymakers' religious beliefs, including the Israeli-Palestinian conflict (Oldmixon, Rosenson and Wald, 2005), reproductive policy (Yamane and Oldmixon, 2006), school prayer (Oldmixon, 2005), and sexual-minority rights (Haider-Markel, 2001; Oldmixon and Calfano, 2007).

<sup>18</sup>The rise of the Christian Right, beginning in the 1960s, aligned Christian conservatism with the Republican party (Dowland, 2015). By 2000, 75-80% of Evangelical Protestants in the U.S. House of Representatives affiliated as Republican, up from roughly 40% between 1973 and 1990 (Newman et al., 2016, 6).

<sup>19</sup>Republican candidates for the U.S. presidency between 1976 and 2004 employed religious rhetoric far more frequently than did their Democratic contenders (Marietta, 2009). Jewish Senators in the 109th-112th Congresses studied here divided evenly among the Democratic and Republican parties.

litical ideologies (Norris and Inglehart, 2011, 211), (Yamane and Oldmixon, 2006).<sup>20</sup>

Given these religious influences on legislative behavior, swaths of American voters expect their candidates to publicly exhibit faith by invoking religious rhetoric on the campaign trail Coe and Chapp (2017).<sup>21</sup> Candidates highlight (often authentic) religious beliefs in a calculated, deliberate, and partisan way to connect with religiously inclined American voters (Domke and Coe, 2008, 3-10) and to demonstrate their moral superiority over their opponents (Glenn, 2010). While this is easier when candidates can “narrowcast” their message to voters with whom they are religiously compatible (Coe and Chapp, 2017), candidates must avoid alienating a religiously diverse electorate or non-religious voters (Albertson, 2015; McLaughlin and Wise, 2014; Weber and Thornton, 2012).<sup>22</sup> The power of this “God Strategy” is contingent on candidate identity<sup>23</sup> but rests in the strength and meaning of religious messages, which shape individual voters’ attitudes, cue political ideological leanings, and activate support among constituents (Newman et al., 2016; Weber and Thornton, 2012).

We contribute to this research by examining three currently understudied dynamics. First, research on religious rhetoric focuses almost exclusively on campaign strategies, in which candidates may shift their rhetoric according to the religious composition of their audiences (Coe and Chapp, 2017). However, congressional discourse also shapes public attitudes and policy outcomes (Leep, 2010). Yet we know relatively little about how officials, *once elected*, publicly and broadly invoke religious rhetoric.<sup>24</sup> Our data enable us to examine elected officials’ religious rhetoric in their communication outreach to their entire constituency.

Second, this existing research emphasizes how a politician’s specific religious identity shapes her electoral strategies and policymaking agendas. It tends to focus on fixed religious identities (e.g., Muslim, Catholic Democrat, Protestant Evangelical)—which can introduce little variation in public opinion outcomes<sup>25</sup>—or private religious practices (Guth, 2014). These fixed measures are unable to sufficiently quantify the ways in which congressmembers publicly perform and invoke their religiosity. Indeed, “[G]iven the entry of religion into political debates issuing in effective policies, and the passionate commitments these debates engender, it makes

<sup>20</sup>For example, while Catholic Democrats tend to prioritize Catholic Social Teachings’ commitments to alleviating public social injustices, Republican Catholics tend to focus on private morality issues like abortion (Oldmixon and Hudson, 2008).

<sup>21</sup>Religion’s role in American presidential campaigns have shifted dramatically since the 1960s, when voters disregarded religious criterion for office (Balmer, 2008). Today, candidates find it necessary to disclose and publicize their religious commitments (Hogue, 2012). Accordingly, conservative religious pundits relayed a popular sentiment that, to be successful, American politicians must at least claim a religious affiliation: “All politicians, Democrats and Republicans alike, love God. Or, more accurately, they love to use God to baptize their political agendas. In the Congressional Directory... no one is an atheist... You never know when it might help you to be religious” (Thomas and Dobson, 1999, 83).

<sup>22</sup>This strategic balance between in- and out-group candidate-voter relationships mirrors Collingwood (2012); Collingwood, Barreto and Garcia-Rios (2014)’s analysis of cross-racial mobilization strategies. Calfano and Djupe (2009) find that candidate appeals to evangelical Protestants do not dissuade mainline Protestant or Catholic voters.

<sup>23</sup>Christian male candidates courting white Republican Evangelical voters disproportionately benefit from this religious-appeals strategy (Calfano and Djupe, 2009). Voters suspect Muslim candidates who lack sizable Muslim constituencies of sympathizing with Muslim terrorists (Braman and Sinno, 2009). Meanwhile, female candidates’ use of religious appeals can reinforce negative gender stereotypes and reduce voter support (Calfano and Djupe, 2011). Furthermore, female politicians find that functioning “within the constraints of a highly gendered religious domain” often dis-empowers their personal sense of agency or efficacy (Calhoun-Brown, 2010, 244).

<sup>24</sup>For example, Wolfe and Katznelson (2010)’s edited volume offers an introduction to American political science’s research on religion. It revealingly provides “no account of the language [elected officials] use to communicate with the public” (Djupe, 2013).

<sup>25</sup>Some evidence suggests that a Christian politician’s specific affiliation has no differential effect on public opinion among constituents with varying religious identities (McKeown and Carlson, 1987).



little sense to measure the social significance of religion only in terms of such indices as church attendance” (Asad, 2003, 182). As a result, researchers lack a comprehensive understanding of the extent to which congressmembers across religious affiliations use religious signaling in their constituent communication strategies. Our measure of public-facing religiosity seeks to capture congressmembers’ more socially formative public religious rhetoric.

Third, the research that has interrogated congressmembers’ religious rhetoric predominately focus on highly partisan issues, like abortion (Marchetti and O’Connell, 2017), in which religious and partisan motivations are not easily distinguishable.<sup>26</sup> However, candidates’ religious rhetoric is frequently more encompassing than issue-specific approaches acknowledge (Chapp, 2012). Thus, this research leaves unanswered how elected officials, across parties, political ideologies, or religious affiliations deploy religious rhetoric. By dislodging our analysis from polarizing issues or partisan politics, we gain insights about the broader dispositional contributions religious ideas make to American politics.

### 3 Theories and Hypotheses

We examine the dispositional attitudes most strongly associated with U.S. senators’ use of religious rhetoric in their public-facing statements. We identify two distinct dispositions (optimism and anxiety) which elected officials may understand and conceptualize within the context of a religious framework. Although these dispositions are not mutually exclusive, we seek to identify which (if either) is most predictive of religious rhetoric.

#### Optimism

A common mis-perception suggests that American religious leaders invoke the threat of insecurity and hell to encourage belief in God and compel behavioral compliance. To the contrary, survey evidence indicates that Americans believe in a “more avuncular than angry” God: “Ever an optimistic people, Americans are more likely to envision heaven than hell... Americans believe in a God who is loving and not very judgmental. Sixty-two percent say that they feel God’s *love* in their life, while only 39 percent say they feel God’s *judgment* with a similarly high frequency in daily life (Putnam and Campbell, 2012, 8). Accordingly, worship services tend to be optimistic and affirming (Wolfe, 2003). Religious circles that emphasize prosperity theology (Bowler, 2018) add an additional layer of optimistic aspiration. Prosperity theology emphasizes God’s causal agency and propensity to distribute favorable outcomes to strong believers: “God grants material prosperity, good health, or relief from sickness to those who have enough faith” (Scott and Hyun, 2012, 738).<sup>27</sup>

If senators adopt this optimistic dispositional application of religion to social circumstances, we expect them to invoke religious references during times of optimism and perceived well-being. Doing so would allow senators to signal peace and prosperity and to engender hope, gratitude, and aspiration among their constituents. To lend support for this hypothesis, we would expect the following: When senators’ constituent communications present an overall optimistic disposition, they will be more likely to exhibit theological gratitude and express optimistic circumstances within the context of their religious beliefs.

<sup>26</sup>At their foundation, issue frames are primarily partisan (Arbour, 2014).

<sup>27</sup>Prosperity gospel is particularly prominent among those with lower levels of education and income (Lugo et al., 2006; Scott and Hyun, 2012).

**H1: Optimism:** Proportionally higher levels of optimistic rhetoric present in senators’ official website text (at the Senator-year level) will yield proportionally higher levels of religious references.

## Anxiety and Insecurity

Alternatively, religion may serve as a source of comfort and stability to communities facing material, economic, or other forms of human insecurity.<sup>28</sup> Countries and population subsets experiencing higher levels of poverty or insecurity exhibit more religiosity than communities who are more secure Norris and Inglehart (2011).<sup>29</sup> Poorer American communities are notably more religious than their wealthier counterparts Norris and Inglehart (2011, 108).<sup>30</sup> Religion may also help people understand and cope with traumatic events. Evidence suggests that even some “secular people turn to religion at times of natural crisis”: religious faith increased among those affected by a 2011 earthquake in New Zealand, despite overall declines in religiosity (Sibley and Bulbulia, 2012). Similarly, high percentages of American respondents turned to religion in the immediate aftermath of the September 11, 2001, terrorist attacks in the United States (Schuster et al., 2001).<sup>31</sup>

If senators treat religion as a mechanism of support amid turmoil, we expect them to invoke religious references during times of anxiety and perceived instability. Doing so would allow senators to elevate transcendental solutions to materially insurmountable dilemmas; catalyze commonly shared religious sentiments; and even identify, codify, or create a perceived common enemy.<sup>32</sup> If perceptions of anxiety and insecurity encourage senators to use religious rhetoric, we would expect the following: When senators’ constituent communications highlight social anxiety and/or security-related threats, they will be more likely to elevate religion as a promise of security or comfort amid turmoil.

**H2: Anxiety and Insecurity:** Proportionally higher levels of anxiety- or security-related rhetoric present in senators’ official website text (at the Senator-year level) will yield proportionally higher levels of religious references.

## 4 Data and Variables

### Congressional Website Archives

Scholars of the U.S. Congress gained valuable insights about members’ policy priorities and constituent outreach strategies by analyzing congressmembers’ floor speeches and press releases.

<sup>28</sup>Some may interpret insecurity or traumatic events as evidence of evil or as divine retribution for sinful behavior (Sohrabzadeh et al 2017).

<sup>29</sup>“The importance of religiosity persists most strongly among vulnerable populations, especially those living in poorer nations, facing personal survival-threatening risks. We argue that feelings of vulnerability to physical, societal, and personal risks are key factor driving religiosity” (Norris and Inglehart, 2011, 4-5).

<sup>30</sup>This mirrors the sentiment President Barack Obama articulated during his 2008 presidential campaign when he said that Americans facing economic uncertainty “cling to guns or religion or antipathy toward people who aren’t like them or anti-immigrant sentiment or anti-trade sentiment as a way to explain their frustrations” (Pilkington, 2008).

<sup>31</sup>In contradistinction to this theory, Eisenstein and Clark (2014, 2017) demonstrate little-to-no direct relationships between psychological security and religious belief, belonging, and behavior.

<sup>32</sup>Political elites’ public rhetoric shape and strengthen social anxieties and public perceptions of insecurity. By framing an issue as a national security threat, they cultivate fear and give the issue greater political salience. They anoint the issue with a sense of urgency, move it beyond the realm of normal politics, and justify suspending “normal” rules and procedures (Lausten and Wæver, 2000; Phillips, 2007; van Rythoven, 2015; Williams, 2003).



However, these alone yield an incomplete analysis of senators’ communication strategies. Official congressional websites—which provide senators a controlled, direct venue for communicating with their constituents—more comprehensively represent senators’ public-facing communication. They may contain a variety of content, including the senator’s biographical information; policy priorities; committee work and leadership roles; constituent services; in-district events; and archives of floor speeches, press releases, and published opinion pieces.<sup>33</sup>

In this article, we analyze parsed Internet-archived text captures of U.S. senators’ official congressional websites during four election years: 2006, 2008, 2010, and 2012.<sup>34</sup> This includes material that has been replaced or otherwise deleted. The Internet Archive organization contracts with the Library of Congress during election years to target, capture, and curate public web content from official U.S. government websites (URLs with the “.gov” domain).<sup>35</sup> This targeted web-crawling approach dramatically increases the quality of this dataset, thus providing the most extensive available record of congressional website content.

For each U.S. senator in the 109-112th Congresses, we create a regular expression to identify her URL “root” (e.g., `murray.senate.gov` for Sen. Patty Murray, D-WA) and collect all Internet Archive website captures within that domain. This data excludes text from senators’ campaign websites, which are distinct from their official congressional websites and outside the government domain. We aggregate the text found on each senator’s pages by election year.<sup>36</sup> To avoid duplication, we limit text analysis to data from senators’ original website captures and all additional content added since the most recent capture.<sup>37</sup> This yields an analysis of text from 141 unique senators<sup>38</sup> among 90 terabytes of “.gov” archived content. To our knowledge, ours is the first social science analysis of this truly “big” congressional website text data source to answer a specific social science question.

## Word-List Variables

We use the Linguistic Inquiry and Word Count (LIWC) dictionary to construct outcome and main explanatory variables. The LIWC software provides pre-established, equivalent word lists, each categorized according to a specific concept or sentiment (Pennebaker et al 2015).<sup>39</sup> The

<sup>33</sup>Websites generally exclude text from senators’ social media activity.

<sup>34</sup>When web archive (known as WARC or ARC) files are collected, they include a webpage’s original HTML infrastructure alongside the text appearing on the page itself. This HTML language complicates textual analysis and is not relevant to our research. Our analysis therefore uses a version of each WARC/ARC file that includes only the text appearing on a given page (text which is “parsed” from the HTML infrastructure).

<sup>35</sup>The current Internet Archive collection ([https://archive.org/details/additional\\_collections](https://archive.org/details/additional_collections)) contains more than 450 billion webpage “captures” (downloads of URL-linked pages and metadata). Of these, approximately 1.1 billion page captures are from the U.S. government’s “.gov” domain between 1996 and 2013. Each “.gov” webpage archive includes its link data (the page’s URL and every other URL or hyperlink found on the page), the parsed text of the page, and the full data content of the page (including HTML-markup language, images, and video files). The captures are house on a Hadoop distributed computing cluster run by SAP Cloud Platform Computing (<https://cloudplatform.sap.com>).

<sup>36</sup>This yearly measure avoids the concerns about how website material was scraped that would be introduced in a daily, weekly, or monthly measure.

<sup>37</sup>One disadvantage of this approach is that it excludes text that was added to a senator’s websites prior to 2006 but remains on her website during a given year of our analysis. To ensure this omission is innocuous, we run a separate, robustness-check analysis among all captures from all years.

<sup>38</sup>Of these, only one—Sen. Bennet, D-CO—names no religious affiliation.

<sup>39</sup>This dictionary method takes a list of words or strings (“explicit mentions” of a given concept) and searches and retrieves exact matches to that list in text data. As a whole, each LIWC list represents a specific concept (e.g., religiosity) or sentiment (e.g., positive emotion or anxiety). The resulting word-list units can be used, as we do here, to compare the prevalence of concepts or sentiments in a given set of text data.

main variables in our analysis represent a proportion of words in a senator’s captured website material in a given year that LIWC categorizes as religious, anxiety-related, or optimistic.

Our outcome variable is a senator’s yearly proportional use of religious terms on her website.<sup>40</sup> This measure of public-facing religiosity represents a senator’s deliberate, public invocation of religious rhetoric in her constituent communication. It correlates highly with an alternative measure, based on Chapp (2012)’s list of religious terms used during the 2012 U.S. presidential campaigns (Appendix); this reinforces the validity of our LIWC-based measure. To our knowledge, this is the first quantitative, comparable, temporally sensitive measure of Congressmembers’ public-facing religiosity.

Our main explanatory variables are a senator’s yearly proportional uses of LIWC’s optimistic terms (H1) and anxiety-related terms (H2). Each resulting measure represents a senator’s deliberate, public invocation of each respective disposition in her public-facing communications. We add an additional “security-related” measure, based on the U.S. Department of Homeland Security (DHS)’s monitoring word list.<sup>41</sup> No comparable government-issued list of optimistic terms are available. To further explore the relationship between a senator’s use of anxiety terms and religious rhetoric (which yields the strongest association), we include measures for real anxiety-inducing events occurring in a given year: state-level Federal Emergency Management Agency (FEMA) emergency declarations,<sup>42</sup> reported state-level terrorist attacks, and international terrorist attacks.<sup>43</sup> Figure 1 depicts the religious, optimistic, anxiety-related, and security-related terms that most commonly appear on senators’ websites. Appendix provides complete word lists.

<sup>40</sup>We added a few key religion-related terms to LIWC’s base-category list and omitted a few others that corresponded with senator names (e.g., Bishop). See Appendix for the complete original and amended lists.

<sup>41</sup>Due to the temporal cost of managing the 90-terabyte “.gov” data cluster, we only include unigrams of the DHS iteration. We exclude terms likely to produce false-positive results (e.g., state or country names). DHS wordlist available at: <https://gist.github.com/jm3/2815378>.

<sup>42</sup><https://www.fema.gov/disasters/state-tribal-government>

<sup>43</sup><https://www.start.umd.edu/gtd/>

Figure 1: Some caption

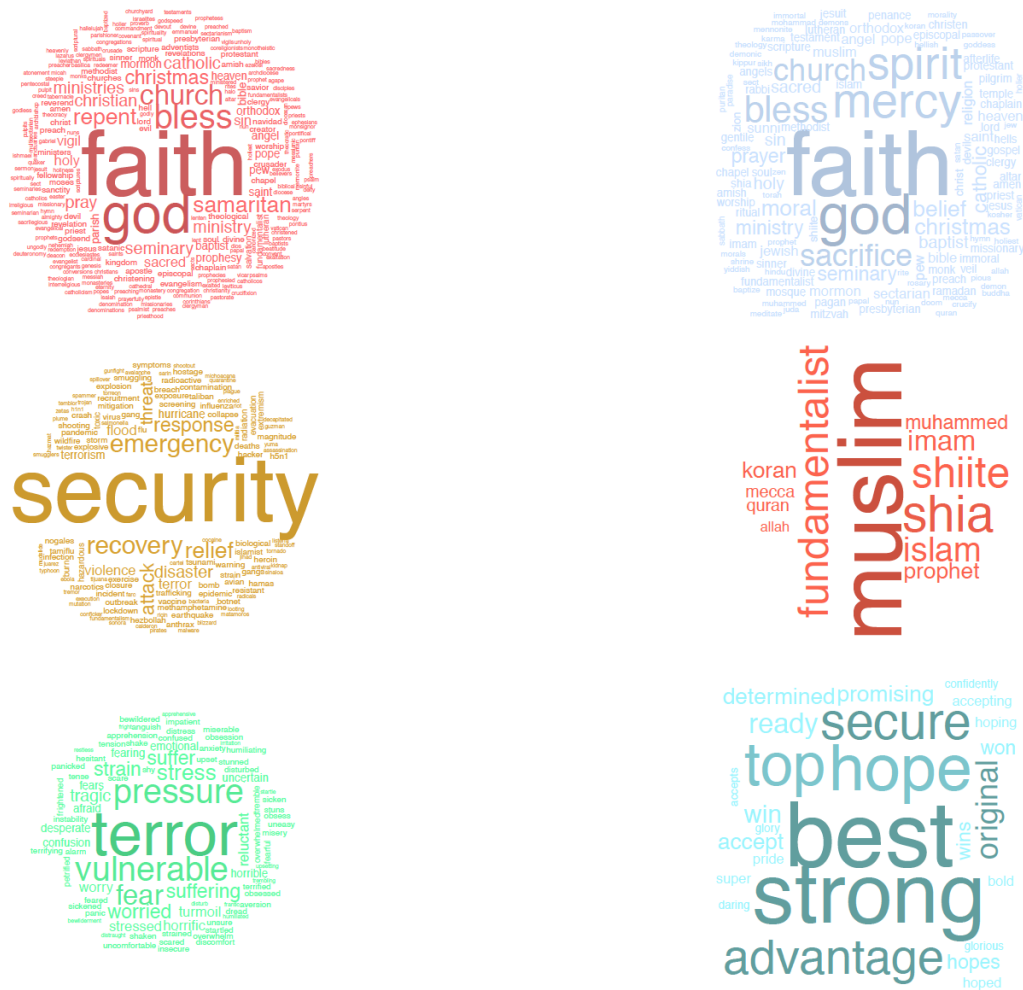


Figure 2: Terms from each word list most commonly appearing on senators' websites. The size of each word corresponds to its relative frequency. *Horizontally from top left:* LIWC religious terms, Chapp (2012) religious terms, DHS security terms, religious terms omitted for their potential overlap with DHS terms, LIWC anxiety terms, and LIWC positive/optimism terms.

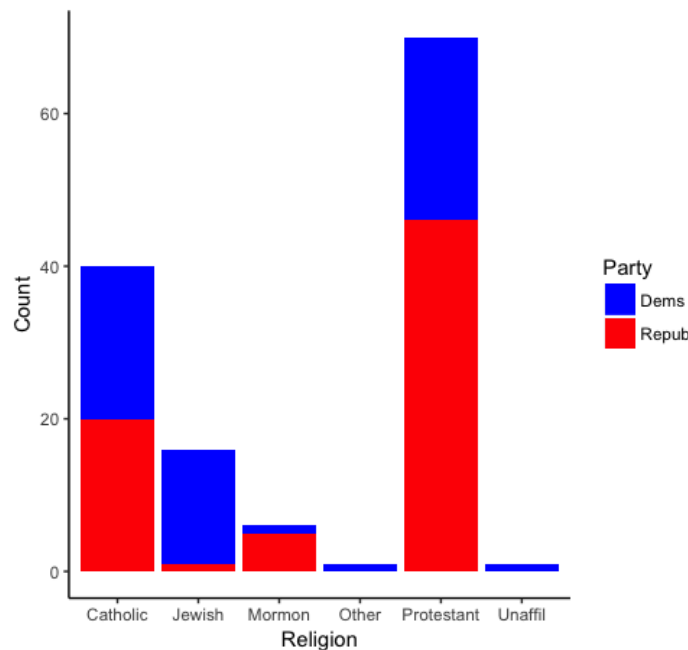
## Senator- and State-Level Controls

We include several demographic factors likely to shape the frequency with which senators use religious rhetoric. Because virtually all senators in our dataset claim a religious affiliation and the considerable majority identify as Christian, it makes little sense to control for religious affiliation. Similarly, Christian denominational specifications can mask such variation as to provide limited meaningful information.<sup>44</sup> However, since Evangelical Protestants—for whom public

<sup>44</sup>For example, Protestantism can represent both a socially progressive, theological interpretivist, and publicly secular mainline Lutheran senator, on one hand, and a socially conservative, theologically dogmatic, and publicly evangelizing Protestant evangelical senator, on the other. Indeed, empirical analyses demonstrate the important differences between specific religious traditions (e.g., between white and Latino Catholics or between black,

evangelizing is often considered paramount—are increasingly associated with conservative political ideologies, we control for political ideology to capture much of these religio-political dynamics.<sup>45</sup> See Figure 3 for senators’ reported religious affiliations by party. We include binary controls to indicate gender, which can mediate religious rhetoric’s political advantage (Calfano and Djupe, 2011), and the year each senator ran for re-election. Finally, we control for state-wide religiosity (the percentage who self-report as “highly religious”<sup>46</sup>) and conservatism (the percent who identify as “very conservative”). Drawing on existing research, we expect that adopting a prominent public-facing religiosity may be a particularly advantageous “God Strategy” to conservative senators, those with highly religious and/or very conservative constituencies,<sup>47</sup> senators up for re-election, and male senators. Appendix Table 2 displays descriptive statistics of key variables.

Figure 3: Senator Religious Affiliations by Party (2006-2012)



## Data Limitations

The Internet Archive gathers and archives website content by programming an automated software program (a “bot”) to select, capture, and sequentially crawl from a seed website URL to all links embedded within that and all subsequently linked websites. It captures all web content (in a WARC or ARC file) along the way. This approach yields unavoidably incomplete and unrepresentative data. The seed URLs the Internet Archive uses are not selected randomly,

mainline, and evangelical Protestants) (Newman et al., 2016, 5), (Guth et al., 2006, 225-26).

<sup>45</sup>Researchers regularly control for political ideology in lieu of partisan identification. Political ideology covaries almost perfectly with political-party identification and offers more variation than party-ID binary variables. We therefore follow this convention here.

<sup>46</sup>This strongly correlates with the percentage of people who report praying regularly (<http://www.pewresearch.org/fact-tank/2016/02/29/how-religious-is-your-state/?state=alabama>).

<sup>47</sup>Although the size of a politician’s religious constituency does not always shape her religious rhetoric (Gin, 2012).

they differ over time, and they are finite (relative to the unknown expanse of available web content<sup>48</sup>). Furthermore, web content changes constantly as data is being captured.<sup>49</sup>

Nevertheless, we reasonably expect that the exceptional quality of Internet-archived congressional website data meets the high standard required for social science analysis. First, the Internet Archive uses the best available methods for archiving the arguably limitless universe of Internet data.<sup>50</sup> Its crawling technology has also dramatically improved over time. For example, Internet Archive software captured White House website content only three times in 1997. It had crawled `whitehouse.gov` at least once a week by 2008 and at least once a day by 2012. (see Figure 4).

Figure 4: Frequency of `whitehouse.gov` crawls



Image from the Internet Archive’s “Wayback Machine.” Notably, White House captures in 2001 temporarily spiked as more web content linked to `whitehouse.gov` in the three months following the September 11 terrorist attacks.

Second, the Library of Congress contracts with Internet Archive to target congressional

<sup>48</sup><http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html>

<sup>49</sup>The Internet Archive also occasionally deletes content at the website owner’s request and does not access secure material on the “dark web.” Neither of these elements apply to congressional websites.

<sup>50</sup>Google and other major search firms use this method.

websites as seed URLs during and immediately after each election cycle (November through January). Together with ever-improving technology, this contracted target has dramatically increased the amount of government Internet material captured and archived.<sup>51</sup> The resulting data are among the most reliably and comprehensively representative in the Internet Archive. By aggregating our observations to the election-year level, we reasonably infer that our dataset approaches containing the universe of congressional text.

## 5 Models and Results

We estimate all models with beta regression. This allows us to appropriately analyze the unique proportional structure of our outcome of interest (public-facing religiosity), in which most values fall close to the lower bound.<sup>52</sup> The formula for beta regression is given as:

$$g(\mu_i) = x_i^T \beta = \eta_i, \quad (1)$$

where  $\beta = (\beta_1, \dots, \beta_k)^T$  represents regression parameters (unknown),  $\eta_i$  houses a linear predictor, and  $g$  represents a link function (Cribari-Neto and Zeileis, 2009, 3). In lieu of including senator-fixed effects (which are incompatible with beta regression), we include a lagged outcome variable. This controls for the fact that senators who used religious rhetoric in the past may be more likely to use it in the future.

We adopt a three-stage analysis. First, we externally ground our outcome variable of interest—senators’ yearly proportional use of religious terms—with data from U.S. House of Representatives’ websites to demonstrate its validity as a measure of public-facing religiosity. Second, we analyze the relationship between senators’ public dispositions (of optimism or anxiety) and their use of religious terms. Finally, we conduct robustness checks to assess the strength of our results and address an endogeneity concern about structural similarities across rhetorical content.

### Grounding Public-Facing Religiosity (U.S. House of Representatives)

If a congressman’s proportional use of religious terms is a meaningful measure of her public-facing religiosity, we expect it would correlate highly with her actual personal religious activity. Although typically not available, Guth (2014) manually tracked the religious activity of the members of the 112th U.S. House of Representatives (2012) and assigned each representative a religious-activity score.<sup>53</sup> Beta regressions demonstrate that Guth’s measure of personal religious activity has a positive, statistically significant ( $p < .001$ ) relationship with representatives’ proportional yearly use of LIWC-identified religious terms at the bivariate and multivariate levels.<sup>54</sup> This validates our measure (of proportional use of religious terms) as a meaningful proxy for public-facing religiosity.

Crucially, representatives’ demographics or observed religious activities have no statistical relationship with any other form of rhetoric analyzed here. Personal religious activity predicts representatives’ proportional use of neither dispositional (optimistic or anxiety-related terms)

<sup>51</sup>The number of unique government website captures nearly triples from 60.2 million in 2003 to 173.5 million during the 2004 election year. In election year 2010, the Internet Archive collected 547.1 million unique “.gov” URLs. This yields a near-complete crawl during election years.

<sup>52</sup>Because beta regressions do not allow true zero values, we increase any zero values to .0000001.

<sup>53</sup>0=no observable involvement, 1=membership in a congregation without observable participation, 2=regular religious-service attendance, 3=leadership participation in congregation.

<sup>54</sup>The same is true of Chapp (2012)’s list of religious terms.



nor DHS security-related terms (Appendix A). This rules out concerns that our models capture a spurious correlation between religiosity and rhetoric.<sup>55</sup> By grounding our measure of public-facing religiosity in Guth’s measure of representatives’ private religious activity, we demonstrate its external validity and reliability. Our proposed measure of public-facing religiosity offers the most socially relevant, time-variant, and broadly attainable approach to measuring and analyzing elected officials’ religiosity.

### Analyzing Predictors of Public-Facing Religiosity (U.S. Senate)

First, models confirm our expectations about the relationships between senator- and state-level demographics, on one hand, and senators’ public-facing religiosity, on the other. Senators who are male and/or more conservative use more religious rhetoric than their peers. Those who represent highly religious constituents also use more religious rhetoric, although the significance of this relationship varies based on model specification (significance disappears in models that include a lagged outcome variable; see Appendix).

Second, we find a strong, statistically significant, positive relationship between both optimism ( $p < 0.001$ ) and anxiety ( $p < 0.0001$ ) and public-facing religiosity. DHS security-related terms, on the other hand, are only associated with religious rhetoric among models that exclude a lagged outcome variable.<sup>56</sup> Senators’ past uses of both anxiety- and security-related rhetoric are the strongest predictors of their future use of those terms (autocorrelation). Most notable in these results is the relationship between a senator’s invocation of anxiety and religious terms. The substantive effect of this relationship is more than double that of the relationship between optimism and religious rhetoric. Senators who publicly display anxious dispositions are likely to frequently invoke public-facing religiosity.

Next, we explore how real anxiety-inducing events influence senators’ dispositional rhetoric. Unexpectedly, no models suggest a relationship between real events (e.g., state-level FEMA declarations, state-level terrorist attacks, or international terrorist attacks) and a senator’s use of anxiety-related words. There is, however, a stable, negative, statistically significant relationship between optimistic terms and terrorist attacks. Although senators do not appear to use anxiety- or security-related terms in relation to real events, they are less likely to display optimistic dispositions when their states have experienced terrorist events.

Despite the strong, stable relationship between senators’ anxiety dispositions and public-facing religiosity, we surprisingly find no relationship between real anxiety-inducing events and senators’ religious rhetoric. Real natural disasters or security threats predict neither anxiety-related dispositions nor religious rhetoric among senators’ websites. It appears that senators display anxiety-related dispositions and religious rhetoric concomitantly but *independently* of state disasters or security threats. Senators may be unknowingly, intentionally, or even strategically promulgating unfounded anxious dispositions among their constituents and then offering their constituents transcendental comforts to assuage those anxieties.

### Robustness Checks

We conduct several robustness checks to evaluate the stability of our findings. First, we replicate our analyses using linear regression in order to include senator-fixed effects. This model makes

<sup>55</sup> While we are forced to remove some representatives from our analysis (do to missing data in ours and/or Guth’s datasets), a Welch Two Sample t-test demonstrates no substantive distance between the representatives in the Guth study and those included ours ( $t = -0.21357$ ,  $df = 834.21$ ,  $p\text{-value} = 0.8309$ ).

<sup>56</sup> The same is true among the positive relationship between DHS terms and real events.

Table 1

	<i>Dependent variable:</i>			
	Frequency Religious Terms (LWIC)			
	(1)	(2)	(3)	(4)
Lagged DV	39.270 (42.995)	1.593 (44.714)	15.034 (42.661)	12.165 (42.715)
Freq Anxiety	507.212*** (121.214)		499.949*** (120.357)	500.792*** (120.165)
Freq DHS	22.976 (27.117)		35.548 (26.021)	32.304 (26.180)
Freq Opt	199.195*** (59.058)		180.928*** (61.666)	176.336*** (61.817)
FEMA Declarations (State)				0.0003 (0.001)
Terrorist Attacks (State)				−0.056 (0.056)
Conservatism (DW1) Score (Senator)		0.327** (0.131)	0.323*** (0.123)	0.321*** (0.123)
Female (Senator)		−0.273** (0.130)	−0.212* (0.126)	−0.216* (0.126)
Very Conservative (State)		−0.324 (0.392)	−0.403 (0.377)	−0.481 (0.382)
Very Religious (State)		0.960 (0.586)	0.852 (0.564)	0.904 (0.575)
Up for Election (Senator)		−0.025 (0.089)	0.026 (0.083)	0.023 (0.083)
Constant	−8.198*** (0.178)	−6.637*** (1.172)	−7.260*** (1.121)	−6.991*** (1.137)
Observations	238	238	238	238
R <sup>2</sup>	0.124	0.047	0.152	0.157
Log Likelihood	1,523.388	1,512.086	1,533.312	1,533.901

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

problematic assumptions about the structure of our data but nevertheless yields results comparable to our main beta-regression models. Second, we reestimate our models using Chapp's (rather than LIWC's) list of religious terms as the basis for measuring senators' public-facing religiosity. This yields very similar results.

Finally, we anticipate and address two possible contributors to "false positive" regression results. First, given cognitive associations many Americans incorrectly make between Islam and terrorism, some U.S. senators may invoke select Islam-related LIWC-identified religious terms under the guise of security concerns. This would create a false positive association between anxiety-, security-, and religious-related terms. Second, senator condolences of "thoughts and prayers" amid tragedy and turmoil could yield similar false positive results. We re-estimate our models on measures of public-facing religiosity that individually exclude Islam-related terms (Appendix) and the word "prayer(s)" (Appendix). Neither exclusions alter our main findings. Each of these robustness checks support and strengthen the reliability of our conclusions.

### **Endogeneity Concern**

Speech is very likely to be structurally related to other types of speech, regardless of content. It is therefore unsurprising that our models demonstrate measures of speech to be more strongly related (relative to senator- or state-level variables) to our outcome speech measures. However, the varying coefficients among different types of speech (as explanatory variables), relative to one another, are quite informative. We therefore remain confident in our conclusion that a uniquely meaningful relationship exists between senators' use of anxiety-related terms and religious rhetoric. The evidence that some measures of speech (e.g., DHS terms) do not significantly relate to religious rhetoric lends additional support to our claims. Furthermore, evidence that religious rhetoric, but not optimistic, anxiety-related, or security-related rhetoric, covaries with Congressmembers' personal religious activity (Appendix) further demonstrates that these speech measures represent substantively distinct concepts.

## **6 Conclusions and Contributions**

In this article, we leveraged a truly "big" text-data source—the Internet Archive's 90-terabyte collection of content from the U.S. government's Internet domain—to contribute new insights about U.S. senators' use of religious rhetoric in public discourse. We used the exceptionally high quality election-year congressional website data to do the following: develop a broadly applicable, time-varying measure of congressmembers' public-facing religiosity; identify the demographic characteristics associated with high-frequency religious rhetoric; and examine the dispositional attitudes most likely to inspire senators to increase their public-facing religiosity. Using our externally grounded measure of congressmembers' public-facing religiosity (their websites' yearly proportional use of religious terms), we find that senators who are conservative, male, and/or representing highly religious constituents are most likely to exhibit high levels of public religiosity. Furthermore, senators who publicly display anxious dispositions in a given year (regardless of real anxiety-inducing state-level events) are strongly likely to invoke public-facing religiosity. As part of their broad public profile, senators may be presenting and promulgating anxious dispositions among their constituents and then modeling religious beliefs as a mechanism for comfort to assuage this perceived anxiety. Whether and to what extent senators offer these (largely unfounded) anxious dispositions and transcendental comforts unknowingly, intentionally, or strategically should be the focus of future study.

This research contributes to our understanding about the role religion plays in American legislative politics. Although religion could serve as an explanation for public prosperity and a venue of aspiring optimism, it appears that American politicians are more readily inclined to leverage religion alongside public anxiety. Religious rhetoric may even help policymakers construct, codify and elevate perceptions of national-security threats, particularly as one particular American religion—Islam—is problematically racialized and associated with insecurity. Considering that religious associations are uniquely advantageous to white, male, Christian politicians, other identity groups may face increased racial or gender-based discrimination if they seek to invoke religious rhetoric. Future research should consider whether congressmembers from various identity groups use, or avoid, religious rhetoric in distinct ways.<sup>57</sup>

Using big data, specifically archived web data, to address social science questions will be required if social science is to stay relevant. This is especially true because web archives house *deleted* content. Yet because of the significant barriers to entry required for this analysis,<sup>58</sup> few social science studies have leveraged the potential power of these resources. Thus, this project has two applied methodological components: first, we construct an original, time varying measure of public facing religiosity which other scholars can use to address their own research questions. Second, by constructing a measure and answer a question that has heretofore not been possible because of the preponderance of man hours that would have been required to do so, we model a means of using big data sources in an empirically grounded manor to advance the frontiers of social science.

We have examined whether specific dispositions and religious rhetoric are associated within a senator’s broad communication profile. When a senator couples anxiety-related dispositions in one week with religious rhetoric the next, these two elements contribute to a macro-level political narrative that associates these two concepts. This research thus establishes a new, multi-methodological research agenda. First, textured qualitative work should explore the mechanisms that connect anxiety and religion in political spaces, as well as the outcomes of this anxiety-religion association for congressmembers, constituents, dominant religious groups, and marginalized religions. Second, future research should examine the relational proximity of these associations (e.g., whether and when these associations occur in a given press release). Researchers could conjoin our approach to measuring public-facing religiosity with natural-language processing (NLP) techniques to examine these relational associations. Third, researchers may use similar NLP techniques to identify congressmembers’ associations between religious rhetoric and specific policy areas.

Finally, researchers should examine the future stability of these anxiety-religion associations. Recent outrage against American gun violence has demonstrated a growing intolerance toward policymakers’ offers of religious “thoughts and prayers” alongside policy inaction. Does this demonstrate a public demand that congressmembers dislodge their invocation of religious rhetoric from times of public crisis? The methods and measures we model here will prove crucial for tracking these and other possible shifts in the role religion plays in American politics.

---

<sup>57</sup>In an effort to strengthen American political science’s considerably weak understanding of Islam (Tepe and Demirkaya, 2011), this research agenda should give particular attention to Muslim politicians.

<sup>58</sup>For example, the data collection effort for this project required use of Hadoop distributed computing systems, bash, python, and SQL.

## References

- Alberts, Thomas. 2008. "Virtually Real: Fake Religions and Problems of Authenticity in Religion." *Culture and Religion* 9(2):125–139.
- Albertson, Bethany L. 2015. "Dog-whistle Politics: Multivocal Communication and Religious Appeals." *Political Behavior* 37(1):3–26.
- Arbour, Brian. 2014. "Issue Frame Ownership: The Partisan Roots of Campaign Rhetoric." *Political Communication* 31(4):604–627.
- Asad, Talal. 2003. *Formations of the Secular: Christianity, Islam, mModernity*. Stanford, CA: Stanford University Press.
- Balmer, Randall Herbert. 2008. *God in the White House: A History: How Faith Shaped the Presidency from John F. Kennedy to George W. Bush*. Grand Rapids, MI: Zondervan.
- Blackstone, Bethany and Elizabeth A Oldmixon. 2015. "Discourse and dissonance: religious agendas in the 104th Congress." *Research & Politics* 2(3):2053168015589215.
- Bowler, Kate. 2018. *Blessed: A history of the American prosperity gospel*. Oxford University Press.
- Braman, Eileen and Abdulkader H. Sinno. 2009. "An Experimental Investigation of Causal Attributions for the Political Behavior of Muslim Candidates: Can a Muslim Represent you?" *Politics and Religion* 2(2):247–276.
- Calfano, Brian R. and Paul A. Djupe. 2011. "Not in His Image: The Moderating Effect of Gender on Religious Appeals." *Politics and Religion* 4:338–354.
- Calfano, Brian Robert and Paul A Djupe. 2009. "God talk: Religious cues and electoral support." *Political Research Quarterly* 62(2):329–339.
- Calhoun-Brown, Allison. 2010. This Far by Faith? Religion, Gender, and Efficacy. In *Religion and Democracy in the United States: Danger Or Opportunity?*, ed. Alan Wolfe and Ira Katznelson. Princeton, NJ: Princeton University Press.
- Chapp, Christopher B. 2012. *Religious rhetoric and American politics: The endurance of civil religion in electoral campaigns*. Cornell University Press.
- Chressanthis, George A, Kathie S Gilbert and Paul W Grimes. 1991. "Ideology, constituent interests, and senatorial voting: The case of abortion." *Social Science Quarterly* pp. 588–600.
- Clifford, Scott and Ben Gaskins. 2016. "Trust me, I believe in God: Candidate religiousness as a signal of trustworthiness." *American Politics Research* 44(6):1066–1097.
- Coe, Kevin and Christopher B Chapp. 2017. "Religious rhetoric meets the target audience: Narrowcasting faith in presidential elections." *Communication Monographs* 84(1):110–127.
- Collingwood, Loren. 2012. *The Pursuit of Victory and Incorporation: Elite Strategy, Group Pressure, and Cross-Racial Mobilization*. University of Washington.
- Collingwood, Loren, Matt A Barreto and Sergio I Garcia-Rios. 2014. "Revisiting Latino voting: Cross-racial mobilization in the 2012 election." *Political Research Quarterly* 67(3):632–645.

- Cribari-Neto, Francisco and Achim Zeileis. 2009. "Beta regression in R."
- de Tocqueville, Alexis. 1835. *Democracy in America*.
- Djupe, Paul A. 2013. "Religion and Democracy in the United States: Danger or Opportunity? Edited by Alan Wolfe and Ira Katznelson. New York, NY: Russell Sage Foundation. 2010. 456 pp. 78.50Cloth, 37.50 Paper." *Politics and Religion* 6(2):450–452.
- Domke, David and Kevin Coe. 2008. *The God strategy: How religion became a political weapon in America*. New York: Oxford University Press.
- Dowland, Seth. 2015. *Family Values and the Rise of the Christian Right*. Philadelphia, PA: University of Pennsylvania Press.
- Dreisbach, Daniel L. 2011. "The Bible in the political rhetoric of the American founding." *Politics and Religion* 4(3):401–427.
- Eisenstein, Marie A. and April K. Clark. 2014. "Political Tolerance, Psychological Security, and Religion: Disaggregating the Mediating Influence of Psychological Security." *Politics and Religion* 7(2):287–317.
- Eisenstein, Marie A. and April K. Clark. 2017. "Explaining Differing Democratic Norm Commitment: Rethinking the Religion-Psychological Security-Democratic Norm Support Connection." *Politics and Religion* 10(4):753–785.
- Fastnow, Chris, J Tobin Grant and Thomas J Rudolph. 1999. "Holy roll calls: Religious tradition and voting behavior in the US House." *Social Science Quarterly* pp. 687–701.
- Geertz, Clifford. 2009. Religion as World-View and Ethic. In *Introducing Religion: Readings From the Classic Theorists*, ed. Daniel L. Pals. Oxford University Press.
- Gin, Willie. 2012. "Jesus Q. Politician: Religious rhetoric in the United States, Australia, and Canada." *Politics and Religion* 5(2):317–342.
- Glenn, Gary D. 2010. "Rhetoric and Religion in the 1984 Campaign." *Political Communication* 5(1):1–13.
- Gohmann, Stephan F and Robert L Ohsfeldt. 1994. "Voting in the US House on Abortion Funding Issues." *American Journal of Economics and Sociology* 53(4):455–474.
- Green, John C and James L Guth. 1991. "Religion, representatives, and roll calls." *Legislative Studies Quarterly* pp. 571–584.
- Guth, James L. 2014. "Religion in the American congress: the case of the US House of Representatives, 1953–2003." *Religion, State & Society* 42(2-3):299–313.
- Guth, James L, Lyman A Kellstedt, Corwin E Smidt and John C Green. 2006. "Religious influences in the 2004 presidential election." *Presidential Studies Quarterly* 36(2):223–242.
- Haider-Markel, Donald P. 2001. "Policy Diffusion as a Geographical Expansion of the Scope of Political Conflict: Same-sex Marriage Bans in the 1990s." *State Politics & Policy Quarterly* 1(1):5–26.



- Hogue, Andrew P. 2012. *Stumping God: Reagan, Carter, and the Invention of a Political Faith*. Waco, TX: Baylor University Press.
- Jenkins, Philip. 2011. *The Next Christendom: The Coming of Global Christianity*. New York, NY: Oxford University Press.
- Lausten, Carsten Bagge and Ole Wæver. 2000. "In Defence of Religion: Sacred Referent Objects for Securitization." *Millennium: Journal of International Studies* 29(3):705–739.
- Layman, Geoffrey C. 1999. "'Culture Wars' in the American Party System: Religious and Cultural Change among Partisan Activists Since 1972." *American Politics Quarterly* 27(1):89–121.
- Leep, Matthew Coen. 2010. "The affective production of others: United States policy towards the Israeli–Palestinian conflict." *Cooperation and Conflict* 45(3):331–352.
- Lugo, Luis, Sandra Stencel, John Green, Timothy S Shah, Brian J Grim, Gregory Smith, Robert Ruby, Allison Pond, Andrew Kohut and Paul Taylor. 2006. Spirit and Power: A 10-Country Survey of Pentecostals. In *Pew Forum on Religion and Public Life*.
- Marchetti, Kathleen and David O'Connell. 2017. "Catholic Politicians and the Politics of Abortion Position Taking." *Politics and Religion* p. 1–28.
- Marietta, Morgan. 2009. "The Absolutist Advantage: Sacred Rhetoric in Contemporary Presidential Debate." *Political Communication* 26(4):388–411.
- Martin, Shane. 2009. "The Congressional Representation of Muslim-American Constituents." *Politics and Religion* 2(2):230–246.
- McKeown, Bruce and James M. Carlson. 1987. "An Experimental Study of the Influence of Religious Elites on Public Opinion." *Political Communication* 4(2):93–102.
- McLaughlin, Bryan and David Wise. 2014. "Cueing God: Religious Cues and Voter Support." *Politics and Religion* 7(2):366–394.
- Meier, Kenneth J and Deborah R McFarlane. 1993. "The politics of funding abortion: State responses to the political environment." *American Politics Quarterly* 21(1):81–101.
- Newman, Brian, James L Guth, William Cole, Chris Doran and Edward J Larson. 2016. "Religion and environmental politics in the US House of Representatives." *Environmental Politics* 25(2):289–314.
- Norris, Pippa and Ronald Inglehart. 2011. *Sacred and secular: Religion and politics worldwide*. Cambridge University Press.
- O'Connor, Robert E and Michael B Berkman. 1995. "Religious determinants of state abortion policy." *Social Science Quarterly* pp. 447–459.
- Oldmixon, Elizabeth A, Beth Rosenson and Kenneth D Wald. 2005. "Conflict over Israel: the role of religion, race, party and ideology in the US House of Representatives, 1997–2002." *Terrorism and Political Violence* 17(3):407–426.

- Oldmixon, Elizabeth A. and William Hudson. 2008. "When Church Teachings and Policy Commitments Collide: Perspectives on Catholics in the U.S. House of Representatives." *Politics and Religion* 1(1):113–136.
- Oldmixon, Elizabeth Anne. 2005. *Uncompromising positions: God, sex, and the US House of Representatives*. Georgetown University Press.
- Oldmixon, Elizabeth Anne and Brian Robert Calfano. 2007. "The Religious Dynamics of Decision Making on Gay Rights Issues in the U.S. House of Representatives, 1993–2002." *Journal for the Scientific Study of Religion* 46(1):55–70.
- Page, Benjamin I, Robert Y Shapiro, Paul W Gronke and Robert M Rosenberg. 1984. "Constituency, party, and representation in Congress." *Public Opinion Quarterly* 48(4):741–756.
- Pew Forum on Religion & Public Life. 2003. Religion and Politics: Contention and Consensus. Technical report Pew Research Center Washington, D.C.: .
- Phillips, Nicola. 2007. "The limits of 'securitization': Power, politics and process in US foreign economic policy." *Government and Opposition* 42(2):158–189.
- Pilkington, Ed. 2008. "Obama angers Midwest voters with guns and religion remark." *The Guardian* April 14, 2008. <https://www.theguardian.com/world/2008/apr/14/barackobama.uselections2008>.
- Putnam, Robert D. and David E. Campbell. 2012. *American Grace: How religion divides and unites us*. New York, NY: Simon and Schuster.
- Richardson, James T and Sandie Wightman Fox. 1972. "Religious affiliation as a predictor of voting behavior in abortion reform legislation." *Journal for the Scientific Study of Religion* pp. 347–359.
- Richardson, James T and Sandie Wightman Fox. 1975. "Religion and voting on abortion reform: A follow-up study." *Journal for the Scientific Study of Religion* pp. 159–164.
- Sandstrom, Aleksandra. 2017. Faith on the Hill: The religious composition of the 115th Congress. Technical report Pew Research Center Washington, D.C.: .
- Schechter, David. 2001. "What drives the voting on abortion policy? Investigating partisanship and religion in the state legislative arena." *Women & Politics* 23(4):61–83.
- Schuster, Mark A, Bradley D Stein, Lisa H Jaycox, Rebecca L Collins, Grant N Marshall, Marc N Elliott, Annie J Zhou, David E Kanouse, Janina L Morrison and Sandra H Berry. 2001. "A national survey of stress reactions after the September 11, 2001, terrorist attacks." *New England Journal of Medicine* 345(20):1507–1512.
- Schwarz, Hunter. 2014. "There are states where you technically can't hold public office if you're an atheist." *Washington Post* July 8, 2014. [https://www.washingtonpost.com/blogs/govbeat/wp/2014/07/08/there-are-states-where-you-technically-cant-hold-public-office-if-youre-an-atheist/?utm\\_term=.3fa8ecd34315](https://www.washingtonpost.com/blogs/govbeat/wp/2014/07/08/there-are-states-where-you-technically-cant-hold-public-office-if-youre-an-atheist/?utm_term=.3fa8ecd34315).
- Scott, Schieman and Jung Jong Hyun. 2012. "'Practical Divine Influence': Socioeconomic Status and Belief in the Prosperity Gospel." *Journal for the Scientific Study of Religion* 51(4):738–756.

- Sibley, Chris G and Joseph Bulbulia. 2012. "Faith after an earthquake: A longitudinal study of religion and perceived health before and after the 2011 Christchurch New Zealand earthquake." *PloS one* 7(12).
- Smidt, Corwin, James L Guth and Lyman A Kellstedt. 2017. *The Oxford handbook of religion and American politics*. Oxford University Press.
- Stark, Rodney and Laurence R Iannaccone. 1994. "A supply-side reinterpretation of the" secularization" of Europe." *Journal for the scientific study of religion* pp. 230–252.
- Stark, Rodney and Roger Finke. 2000. *Acts of faith: Explaining the human side of religion*. Univ of California Press.
- Tatalovitch, Raymond and David Schier. 1993. "The persistence of ideological cleavage in voting on abortion legislation in the House of Representatives, 1973-1988." *American Politics Quarterly* 21(1):125–139.
- Tepe, Sultan and Betul Demirkaya. 2011. "(Not) Getting Religion: Has Political Science Lost Sight of Islam?" *Politics and Religion* 4(2):203–228.
- Thomas, Cal and Ed Dobson. 1999. *Blinded by Might: Can the Religious Right Save America?* Grand Rapids, MI: Zondervan Publishing House.
- van Rythoven, Eric. 2015. "Learning to feel, learning to fear? Emotions, imaginaries, and limits in the politics of securitization." *Security Dialogue* 46(5):458–475.
- Waldman, Steven. 2009. *Founding Faith: How Our Founding Fathers Forged a Radical New Approach to Religious Liberty*. Random House Incorporated.
- Weber, Christopher and Matthew Thornton. 2012. "Courting Christians: How Political Candidates Prime Religious Considerations in Campaign Ads." *The Journal of Politics* 74(2):400–413.
- Williams, Michael C. 2003. "Words, Images, Enemies: Securitization and International Politics." *International Studies Quarterly* 47(4):511–531.
- Wing, Nick. 2017. "Here are all the Atheists in Congress." *HuffPost* December 6, 2017. [https://www.huffingtonpost.com/2013/09/19/atheists-in-congress\\_n\\_3944108.html](https://www.huffingtonpost.com/2013/09/19/atheists-in-congress_n_3944108.html).
- Wolfe, Alan. 2003. *The transformation of American religion: How we actually live our faith*. Simon and Schuster.
- Wolfe, Alan and Ira Katznelson. 2010. *Religion and Democracy in the United States: Danger Or Opportunity?* Princeton, NJ: Princeton University Press.
- Yamane, David and Elizabeth A Oldmixon. 2006. "Religion in the legislative arena: Affiliation, salience, advocacy, and public policymaking." *Legislative Studies Quarterly* 31(3):433–460.

## A Appendix: Tables and Figures

EKG to fix - what descriptive stats do we want? This table is from old analysis

Table 2: Descriptive Statistics

Var Name	Min	Max	Mean	SD
Freq Religious	0.000000001	0.002	0.0001	0.0001
Freq Religion No Islam	0.000000001	0.002	0.0001	0.0001
Freq DHS	0.000000001	0.004	0.001	0.0005
State Terror Attacks	0	8	0.28	0.90
FEMA Decs	0	470	30	56
% Very Religious (State)	0.33	0.77	0.54	0.10
Very Conservative (State)	2.83	3.87	3.40	0.18
% Evangelical (State)	2.21	43.97	15.759	11.83
Don't Pray	2.17	4.78	3.16	0.48
Conservatism (Senator)	-0.64	1	0.02	0.43
Jewish Faith (Senator)	0	1	0.12	0.33
Mormon Faith (Senator)	0	1	0.04	0.20
Female (Senator)	0	1	0.16	0.37
Up For Election	0	1	0.33	0.47
Republican (Senator)	0	1	0.47	0.50

## B Appendix: Comparing House Member Religiosity to House Member Use of Religious Rhetoric

Figure 5: Correlation Plot - House of Representatives. Insignificant Correlations are Blank

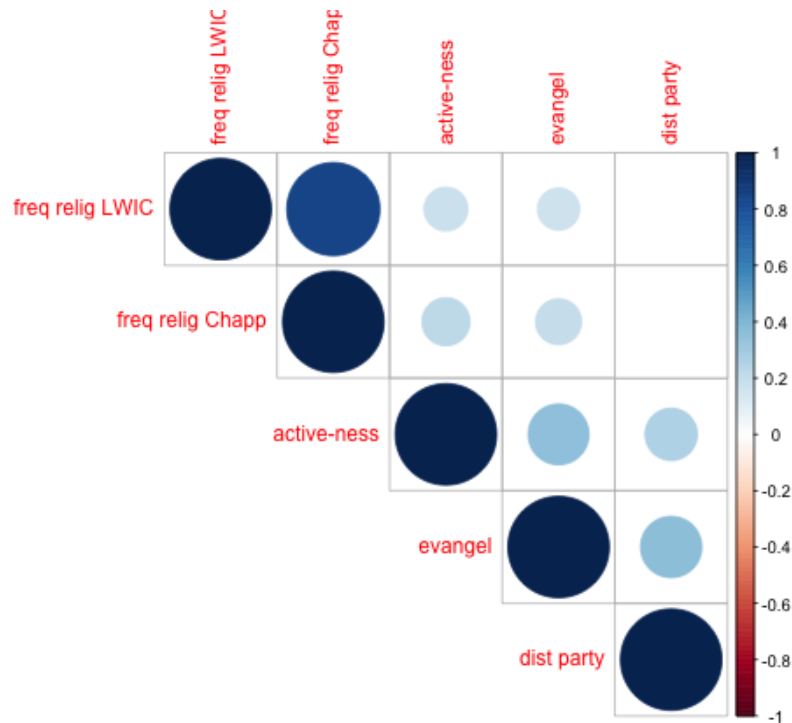


Table 3: House of Representatives: Comparing Measures of Religious Rhetoric

	<i>Dependent variable:</i>			
	LWIC ReligFreq		Chapp ReligFreq	
	(1)	(2)	(3)	(4)
Member Active	0.092*** (0.034)	0.077** (0.037)	0.112*** (0.033)	0.092** (0.036)
Evangelical		0.132 (0.108)		0.155 (0.104)
District Party		0.0004 (0.003)		0.001 (0.003)
Gender M		-0.005 (0.118)		0.015 (0.115)
Constant	-7.418*** (0.077)	-7.420*** (0.127)	-7.578*** (0.075)	-7.596*** (0.124)
Observations	392	392	391	391
R <sup>2</sup>	0.021	0.024	0.038	0.044
Log Likelihood	2,465.513	2,466.398	2,509.145	2,510.696
Model	Beta	Beta	Beta	Beta
Year	2012	2012	2012	2012

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01



Table 4: House of Representatives: Evaluating Non-Religious Rhetoric

	<i>Dependent variable:</i>					
	Freq. DHS		Freq. Opt		Freq. Anx	
	(1)	(2)	(3)	(4)	(5)	(6)
Member Activity	−0.013 (0.032)	−0.002 (0.035)	−0.001 (0.030)	−0.018 (0.033)	0.036 (0.033)	0.038 (0.036)
Evang. (District)		−0.102 (0.106)		0.017 (0.098)		0.028 (0.108)
Party (District)		0.001 (0.003)		0.004 (0.003)		−0.001 (0.003)
Male		−0.112 (0.110)		−0.001 (0.106)		−0.055 (0.115)
Constant	−5.941*** (0.068)	−5.841*** (0.115)	−6.465*** (0.063)	−6.441*** (0.111)	−7.644*** (0.073)	−7.607*** (0.122)
Observations	399	399	399	399	399	399
R <sup>2</sup>	0.0003	0.003	0.00000	0.004	0.002	0.003
Log Likelihood	1,980.793	1,981.733	2,189.634	2,191.027	2,630.407	2,630.688

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

## C Appendix: Additional Regression Results, Senate

Table 5: Bivariate Relationships

	<i>Dependent variable:</i>			
	frequency_religLWIC			
	(1)	(2)	(3)	(4)
Freq. Anx	790.391*** (95.611)			631.333*** (99.648)
Freq. DHS		113.318*** (21.358)		60.306*** (22.107)
Freq. Opt			236.212*** (53.364)	177.385*** (54.149)
Constant	-7.909*** (0.091)	-7.727*** (0.099)	-7.789*** (0.126)	-8.403*** (0.153)
Observations	382	382	382	382
R <sup>2</sup>	0.087	0.039	0.068	0.134
Log Likelihood	2,443.570	2,420.939	2,422.167	2,455.999

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 6: Considering Chapp's Measure of Religious Rhetoric

	<i>Dependent variable:</i>			
	Frequency Chapp			
	(1)	(2)	(3)	(4)
Freq. Relig LWIC (lag)	50.544 (46.378)	17.122 (47.488)	28.367 (46.623)	25.602 (46.709)
FEMA Dec.				0.0003 (0.001)
Terror Attack (State)				-0.051 (0.061)
Freq. Anx	401.839*** (138.013)		373.470*** (137.094)	375.560*** (137.080)
Freq. DHS	38.544 (29.430)		53.680* (27.970)	50.056* (28.193)
Freq. Opt	191.097*** (65.849)		167.668** (68.749)	163.182** (68.951)
Conservatism (DW1)		0.265* (0.141)	0.272** (0.137)	0.270** (0.137)
Female		-0.246* (0.139)	-0.200 (0.138)	-0.205 (0.138)
Very Conservative (% of State)		-0.202 (0.424)	-0.257 (0.419)	-0.329 (0.424)
Very Religious (% of State)		1.300** (0.635)	1.131* (0.626)	1.173* (0.638)
Up For Election		-0.085 (0.097)	-0.034 (0.093)	-0.037 (0.093)
Constant	-8.278*** (0.198)	-7.360*** (1.268)	-7.963*** (1.248)	-7.708*** (1.267)
Observations	238	238	238	238
R <sup>2</sup>	0.111	0.056	0.147	0.151
Log Likelihood	1,534.123	1,529.895	1,543.327	1,543.754

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 7: Evaluating Other Measures of Rhetoric

	<i>Dependent variable:</i>		
	Freq. DHS	Freq. Anx	Freq. Opt
	(1)	(2)	(3)
Freq. DHS_lag1	132.010*** (15.697)		
Freq. Anx_lag1		654.260*** (87.605)	
Freq. Opt_lag1			212.502*** (41.827)
Terror Attack (State)	−0.069* (0.038)	−0.009 (0.041)	−0.120*** (0.040)
globalterrorism	0.00001 (0.00002)	0.00001 (0.00002)	0.00002 (0.00002)
FEMA Dec.	0.001 (0.0004)	0.0001 (0.0005)	0.001 (0.0004)
Conservatism (DW1)	−0.072 (0.085)	0.158* (0.093)	0.093 (0.077)
Female	0.045 (0.075)	0.076 (0.087)	0.019 (0.073)
Very Conservative (% of State)	−0.212 (0.264)	−0.315 (0.297)	0.065 (0.241)
Very Religious	0.390 (0.396)	0.585 (0.440)	0.100 (0.360)
Up For Election	−0.188*** (0.058)	−0.137** (0.064)	−0.112** (0.054)
Constant	−5.654*** (0.814)	−6.971*** (0.925)	−6.980*** (0.739)
Observations	238	238	238
R <sup>2</sup>	0.075	0.080	0.032
Log Likelihood	1,213.083	1,568.897	1,353.369

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 8: No Islam-related Words

	<i>Dependent variable:</i>		
	frequency_reli_Lwicnoislam		
	(1)	(2)	(3)
Freq.Rel_LWIC No Islam (lag 1)	-1.207 (45.411)	35.643 (43.614)	10.052 (43.461)
FEMA Dec.			0.0003 (0.001)
Terror Attack (State)			-0.058 (0.056)
Conservatism (DW1)	0.301** (0.131)		0.289** (0.124)
Female	-0.247* (0.129)		-0.186 (0.126)
Very Conservative (% of State)	-0.244 (0.393)		-0.404 (0.385)
Very Religious	0.996* (0.587)		0.965* (0.580)
Up For Election	-0.032 (0.090)		0.012 (0.084)
Freq. Anx		474.618*** (122.974)	466.892*** (122.200)
Freq. DHS		20.481 (27.289)	27.484 (26.439)
Freq. Opt		201.501*** (59.231)	179.616*** (62.038)
Constant	-6.973*** (1.173)	-8.206*** (0.179)	-7.292*** (1.147)
Observations	238	238	238
R <sup>2</sup>	0.045	0.117	0.150
Log Likelihood	1,521.575	1,531.559	1,541.334

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 9: No Prayer

	<i>Dependent variable:</i>			
	Freq. Relig LWIC No Prayer			
	(1)	(2)	(3)	(4)
Freq. Relig LWIC No Prayer (lag)	41.269 (43.725)	2.840 (45.489)	17.356 (43.286)	14.367 (43.325)
FEMA Dec.				0.0004 (0.001)
Terror Attack (State)				-0.054 (0.055)
Freq. Anx	513.462*** (120.370)		507.040*** (119.508)	507.718*** (119.339)
Freq. DHS	23.302 (26.976)		36.514 (25.872)	33.112 (26.031)
Freq. Opt	199.566*** (58.732)		181.113*** (61.359)	176.495*** (61.519)
Conservatism (DW1)		0.339*** (0.131)	0.332*** (0.123)	0.329*** (0.123)
Female		-0.268** (0.130)	-0.208* (0.125)	-0.213* (0.125)
Very Conservative (% of State)		-0.352 (0.390)	-0.419 (0.375)	-0.497 (0.379)
veryReligious		0.946 (0.584)	0.834 (0.560)	0.880 (0.571)
Up For Election		-0.021 (0.089)	0.028 (0.083)	0.025 (0.083)
Constant	-8.224*** (0.178)	-6.555*** (1.168)	-7.225*** (1.114)	-6.955*** (1.131)
Observations	238	238	238	238
R <sup>2</sup>	0.125	0.047	0.153	0.158
Log Likelihood	1,528.612	1,516.720	1,538.597	1,539.193

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01



## D Appendix: Computational Scripts

The following script flags senator webpages that include one or more mentions of the listed terms and stores a count of those captures. We begin with an overview of the process of running jobs on the cluster, and then provide specific code. For questions, please contact the author (ekgadeuw.edu).

### D.1 Overview

Running scripts on the cluster requires a basic understanding of bash (Unix) shell commands using the Command Line on a home computer (on a Mac, this is the program “Terminal”). For a basic run down of bash commands, see [http://cli.learncodethehardway.org/bash\\_cheat\\_sheet.pdf](http://cli.learncodethehardway.org/bash_cheat_sheet.pdf).

Begin by opening a bash shell on a home desktop, and using an ssh key obtained from Altiscale to log in. Once logged in, you will be on your personal workbench and now have to use a script editor (such as Vi <http://www.catonmat.net/download/bash-vi-editing-mode-cheat-sheet.pdf>). Come up with a name for the script, open the editor, and then either paste or write the desired script in the editor, close and save the file (to your personal workbench on the cluster).

Scripts must be written in Hadoop-accessible languages, such as Apache Pig, Hive, Graph or Oozie. Apache languages are SQL-like, which means if you have experience with SQL, MySQL, SQLite or PostgreSQL (or R or Python), the jump should not be too big. For text processing, Apache Pig is most appropriate, whereas for link analysis, Hive is best. The script below is written in Apache Pig and a manual can be found at <https://pig.apache.org/>. For an example of some scripts written for this cluster, see <https://webarchive.jira.com/wiki/display/Iresearch/IA++GOV+dataset++Altiscale>. May be easiest to it “clone” the “archive analysis” file hosted on GitHub from Vinay Goel <https://github.com/vinaygoel> or three basic scripts from Emily Gade <https://github.com/ekgade/.govDataAnalysis> and use those as a launch point. If you don’t know how to use GitHub, see here: <https://guides.github.com/activities/hello-world/> (it is actually quite straightforward).

Because Apache languages have limited functionality, users may want to write user defined functions in a program like Python. A tutorial about how to do this can be found at [https://help.mortardata.com/technologies/pig/writing\\_python\\_udfs](https://help.mortardata.com/technologies/pig/writing_python_udfs).

Once a script is written, you will want to run it on a segment of the cluster. This requires another set of Unix style Hadoop shell commands (see <http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/FileSystemShell.html>). Users must then specify the file path(s), the desired output directory, and where the script can be found.

### D.2 Getting a Key

As discussed above, this script is run from your workbench on the cluster. To gain access, you will need to set up an SSH “key” with Altiscale (see <http://documentation.altiscale>.

com/configure-ssh-from-mac-linux). Once you have obtained and sent your SSH key to Altiscale, you can log in using any bash shell from your desktop with the command “ssh altiscale”.

### D.3 Locating the Data

The Altiscale cluster houses 9 “buckets” of .GOV data. Each bucket contains hundreds or thousands of Web Archive Files (older version are “ARC” files, newer version are “WARC” files, but they have all the same fields). Each WARC/ARC file contains captures from the same crawl, but it (a) won’t contain all of the captures from a given crawl, and (b) since the crawl is doing a lot of things simultaneously, captures of a single site can be located in different WARC files.

With so much data, there is no simple “table” or directory that can be consulted to locate a specific web page. The best way to find specific pages is to use Hive to query the CDX database. See Vinay Goel’s GitHub for details about how to query CDX: <https://github.com/vinaygoel/archive-analysis/tree/master/hive/cdx>. If a user know exactly what he or she wants (all the captures of the whitehouse.gov mainpage, or all the captures from September 11, 2001), the CDX can tell you where to find them. Otherwise, users will want to query all of the buckets because there is no easy way to learn where results are stored. (Though we advise first testing scripts on a single bucket or WARC file.)

First, use the command line with SSH interface to query the data directories and see which buckets or files to run a job over. This requires the Hadoop syntax to “talk” to the cluster where all the data is stored. The cluster has a user-specific directory where users can store the results of scrapes. A user’s local work bench does not have enough space to save them.

Whenever users “talk” from a user’s local workbench to the main cluster, users need to use ‘hadoop fs -’ and then the bash shell command of interest. For a list of Hadoop-friendly bash shell commands, see: [http://hadoop.apache.org/docs/current1/file\\_system\\_shell.html](http://hadoop.apache.org/docs/current1/file_system_shell.html). For example, the line of code

```
hadoop fs -ls
```

pulls a listing of the files in your personal saved portion of the cluster (in addition to the local workbench, each user has a file directory to save the results). As well,

```
hadoop fs -ls /dataset-derived/gov/parsed/arcs/bucket-2/
```

would draw up all the files in Bucket #2 of the parsed text ARCS directory.

### D.4 Defining Search Terms

Scripts that deal with text are best written in Apache Pig. Hadoop also supports Apache Hive, Giraffe and Spark. To find and collect terms or URLs of interest, users will need to write a script. For example, users might write a script to flag any captures that have a mention of a global warming term, and return the date of the capture, URL, page title, checksum, and the parsed text. This script is saved on your local workbench and needs to have a .pig suffix. Users will need to use some sort of bash editor to write and store the script such as vi (details about how to use vi can be found above). Script is below. The first four lines are defaults and also

set the memory.

Script begins:

```
SET default_parallel 100;

SET mapreduce.map.memory.mb 8192;
SET mapred.max.map.failures.percent 10;
set mapreduce.reduce.memory.mb 16000;
set mapreduce.reduce.java.opts -Xmx8196m

REGISTER lib/ia-porky-jar-with-dependencies.jar;
```

The sequence file loader pulls the files out of the ARC/WARC format and makes them readable. Note, when they were put into the ARC/WARC format, they were run through a HTML parser to remove the HTML boilerplate. However, if the file was not in HTML to begin with, the parser will just produce symbols and this won't fix it. Users will have to deal with those issues separately.

This block allows you to load user defined functions from a Python file:

```
REGISTER 'extrawordsEKGJonJohn.py' USING jython AS myfuncs;
DEFINE FROMJSON org.archive.porky.FromJSON();
DEFINE SequenceFileLoader org.archive.porky.SequenceFileLoader();
DEFINE SURTURL org.archive.porky.SurtUrlKey();
```

When loading data on the command line (instructions below), give the data a name (here \$I\_Parsed\_Data) and make sure to use the same "name" for the data in the command line command. This is a stand-in for the name of the directory or file over which you will run a script.

```
Archive = LOAD '$I_PARSED_DATA' USING SequenceFileLoader() AS
(key:chararray, value:chararray);
```

The below code block says: for each value and key pair, pull out the following fields. Chararray means character array - so a list of characters with no limits on what sort of content may be included in that field. The next line selects the date string. The full format is year, month, day, hour, second. Also note that because Pig is under-written in Java, users need two escape characters in these scripts (whereas only one is needed in Python). Note that "myfuncs.Threat\_countWords" loads the Python UDF (below). If a user has function which selects certain URLs of interest and groups all other URLs as "other", they would run it only on the URL field. And, if a user has a function that collects words of interest and counts them as well as total words, the user should run that through the content field.

```
Archive = FOREACH Archive GENERATE FROMJSON(value) AS m:[];
```

```
Archive = FILTER Archive BY m#'errorMessage' is null;
```

```
ExtractedCounts = FOREACH Archive GENERATE myfuncs.pickURLs(m#'url'),
    m#'url' AS src:chararray,
    SURTURL(m#'url') as surt:chararray,
    REPLACE(m#'digest','sha1:', '') AS checksum:chararray,
    m#'date' as date:chararray,
    myfuncs.Threat_countWords(m#'boiled');
```

In Pig, and the default delimiter is '

n' (new line) but many '

n' appear in text. So one must get rid of all the new lines in the text. This will affect our ability to do text parsing by paragraph, but sentences will still be possible. Code to get rid of the '

n' (new line delimiters) which are causing problems with reading in tables might look something like this:

```
UniqueCaptures = FOREACH UniqueCaptures GENERATE REPLACE(content, '\n', ' ');
```

To get TOTAL number of counts of webpages, rather than simply unique observations, merge with checksum data:

```
Checksum = LOAD '$I_CHECKSUM_DATA' USING PigStorage()
AS (surt:chararray, date:chararray, checksum:chararray);
```

Then join, flatten and output to the directory you listed in the command line:

```
FullCounts = FOREACH CountsJoinChecksum GENERATE
    ExtractedCounts::src as src,
    Checksum::date as date,
    ExtractedCounts::counts as counts,
    ExtractedCounts::URLs as URLs;
```

```
GroupedCounts = GROUP FullCounts BY URLs;
```

```
GroupedCounts = FOREACH GroupedCounts GENERATE
    group AS src,
    FLATTEN(FullCounts);
```

```
GroupedCounts = FOREACH GroupedCounts GENERATE
    src AS src,
    date AS date,
    SUBSTRING(date, 0,4),
    SUBSTRING(date, 4,6),
    URLs AS URLs,
    FLATTEN(counts);
STORE GroupedCounts INTO '$O_DATA_DIR';
```

The UDFs mention here are written in Python and can be seen in at the bottom of this Appendix.

## D.5 Running the Script

To run this script, type the following code into the command line, after having logged in the Altiscale cluster with your ssh key. Users will select the file or bucket they want to run the script over, and type in an “output” directory (this will appear on your home/saved data on the cluster, not on your local workbench). Finally, users need to tell Hadoop which script they want to run. The `$I_PARSED_DATA` was defined as the location of the data to run the script over in the script above. Here we telling the computer that this bucket is the `$I_PARSED_DATA`. Next, one must load the `$CHECKSUM` data, and finally, give the output directory, and the location of your script.

The following should be run all as one line:

```
pig -p I_PARSED_DATA=/dataset-derived/gov/parsed/arcs/bucket-2/  
-p I_CHECKSUM_DATA=/dataset/gov/url-ts-checksum/  
-p O_DATA_DIR=place_where_you_want_the_file_to_end_up  
location_of_your_script/scriptname.pig
```

## D.6 Concatenate Results

Finally, we concatenate results across buckets on the cluster before outputting them.

```
SET default_parallel 20;  
  
SET mapreduce.map.memory.mb 8192;  
SET mapred.max.map.failures.percent 10;  
  
REGISTER lib/ia-porky-jar-with-dependencies.jar;  
  
DEFINE FROMJSON org.archive.porky.FromJSON();  
DEFINE SequenceFileLoader org.archive.porky.SequenceFileLoader();  
  
WordCounts = LOAD '$I_WORD_COUNTS' AS (url1:chararray, timestamp:int,  
year:int, month:int, url:chararray, word:chararray, count:int);  
  
-- note this is the location of the output from the previous script  
  
GroupedCounts = GROUP WordCounts BY (year, month, url, word);  
  
AggregatedCounts = FOREACH GroupedCounts GENERATE  
    group.year AS year, group.month AS month, group.url AS url, group.word AS word,  
    SUM(WordCounts.count) as count;  
  
STORE AggregatedCounts INTO '$O_DATA_DIR';
```

## D.7 Exporting Results

Lastly, to remove results from the cluster users need to open a new Unix shell on their local machine that is NOT logged in to the cluster with their ssh key. Then type the location of the

file they'd like to copy and give it a file path for where they'd like to put it on their desktop. For example:

The following should be run all as one line:

```
scp -r altiscale:~/results_location  
/location_on_your_computer_you_want_to_move_results_to/
```

For additional scripts and for those with programming experience, see Vinay Goel's GitHub at <https://github.com/vinaygoel/archive-analysis>. For stepwise instruction of a wordcount script, see Emily Gade's GitHub at <https://github.com/ekgade/.govDataAnalysis>.

Python UDFs:

```
from collections import defaultdict  
import sys  
import re  
from string import punctuation  
  
@outputSchema("URLs:chararray")  
def pickURLs(url):  
    try:  
        names =set([  
            'tomudall',  
            'wicker',  
            'menendez',  
            'warner',  
            'moran',  
            'voinovich',  
            'webb',  
            'whitehouse',  
            'wyden'])  
  
        regexp = re.compile ('([a-z]+)?(\\.?\\/senate\\.gov\\/?.?)([a-z]+)?')  
        results = []  
  
        result = regexp.search(url)  
        if result is not None:  
            if len(result.group(1)):  
                if result.group(1) in names:  
                    return(result.group(1))  
            if len(result.group(3)):  
                if result.group(3) in names:  
                    return(result.group(3))  
  
    except:  
        pass  
    return 'other'
```

```

# counting words

#define output schema as a "bag" with the word and then the count of the word
@outputSchema('counts:bag{tuple(word:chararray,count:int)}')
def Threat_countWords(content):
    try:
        Threat_Words = set([
            'afterlife',
            'agonstic',
            'alla',
            'allah',
            'altar',
            'amen',
            'amish',
            'angel',
            'upset',
            'vulnerable',
            'worry'])
    except:
        pass
    threat_counts = defaultdict(int)
    threat_counts['total'] = 0

    if not content or not isinstance(content, unicode):
        return [(['total'], 0)]
    splitcontent = content.lower().split()
    threat_counts['total'] = len(splitcontent)
    for word in splitcontent:
        if word in Threat_Words:
            threat_counts[word] += 1

    # Convert counts to bag
    countBag = []
    for word in threat_counts.keys():
        countBag.append( (word, threat_counts[word] ) )
    return countBag

```

## E Appendix: Lists of Terms

### E.1 Religious Words

afterlife, agonstic, alla, allah, altar, amen, amish, angel, angelic, angels, baptist, baptize, belief, bible, biblic, bishop, bless, buddha, catholic, chapel, chaplain, christ, christen, christian, christmas, church, clergy, confess, convents, crucify, demon, demonic, demons, devil, divine, doom, episcopal, evangelic, faith, fundamentalist, gentile, god, goddess, gospel, hashanal, heaven, hell, hellish, hells, hindu, holier, holiest, holy, hymn, imam, immoral, immortal, islam, jesuit, jesus,

jew, jewish, juda, karma, kippur, koran, kosher, lord, lutheran, mecca, meditate, mennonite, mercifull, mercy, methodist, minister, ministry, missionary, mitzvah, mohammad, monastery, monk, moral, morality, morals, mormon, mosque, muhammed, mujahids, muslim, nun, orthodox, pagan, papal, paradise, passover, pastor, penance, pentecost, pew, piet, pilgrim, pious, pious, pope, prayer, preach, presbyterian, priest, prophet, protestant, puritan, quran, rabbi, rabbinica, ramadan, religion, rite, ritual, rosary, sabbath, sacred, sacrifice, saint, salvation, satan, scripture, sect, sectarian, seminary, shia, shiite, shrine, sikh, sin, sinner, soul, spirit, sunni, temple, testament, theology, torah, vatican, veil, worship, yiddish, zen, zion, christian, christianity, hell, monastery, pagans, believer, believers, blessed, bless, wrath, almighty, christ, grace

## **E.2 DHS Words**

anthrax, antiviral, assassination, attack, avalanche, avian, bacteria, biological, blizzard, bomb, botnet, breach, burn, calderon, cartel, closure, cocaine, collapse, conficker, contamination, crash, deaths, decapitated, disaster, earthquake, ebola, emergency, enriched, epidemic, evacuation, execution, exercise, explosion, explosive, exposure, extremism, farc, flood, flu, fundamentalism, gang, gangs, gunfight, guzman, h1n1, h5n1, hacker, hamas, hazardous, hazmat, heroin, hezbollah, hostage, hurricane, incident, infection, influenza, islamist, jihad, juarez, keylogger, kidnap, listeria, lockdown, looting, magnitude, malware, matamoros, methamphetamine, mexicles, michoacana, militia, mitigation, mudslide, mutation, narcos, narcotics, nogales, outbreak, pandemic, pirates, plague, plume, quarantine, radiation, radicals, radioactive, recovery, recruitment, relief, resistant, response, reynose, reyosa, ricin, riot, rootkit, salmonella, sarin, screening, security, shooting, shootout, sinaloa, smugglers, smuggling, sonora, spammer, spillover, stand-off, storm, strain, symptoms, taliban, tamaulipas, tamiflu, temblor, terror, terrorism, threat, tijuana, tornado, torreon, toxic, trafficking, tremor, trojan, tsunami, twister, typhoon, vaccine, violence, virus, warning, wildfire, yuma, zetas

## **E.3 Anxiety Words**

afraid, alarm, anguish, anxiety, apprehension, aversion, bewilderment, confusion, desperate, discomfort, distraught, distress, disturb, dread, emotional, fear, feared, fearing, fears, frantic, fright, hesitant, horrific, horrible, humiliating, impatient, inadequate, insecure, irritation, misery, numerous, obsession, obsess, overwhelm, panic, petrify, pressure, reluctant, restless, saw, scare, shake, shy, sicken, startle, strain, stress, stunned, stuns, tense, tension, terrified, terrifying, terror, tremble, turmoil, uncertain, uncomfortable, uneasy, unsure, upset, vulnerable, worry, fearful, worried, scared, suffer, suffering, need, help, miserable, apprehensive, bewildered, confused, disturbed, fearful, frightened, humiliated, miserable, obsessed, overwhelmed, panicked, petrified, scared, shaken, sickened, startled, strained, stressed, tragic, trembling, instability, upsetting, concerned