Measuring Policy Positioning in U.S. Congressional Elections*

Colin R. Case† Rachel Porter‡

August 30, 2024

Abstract

Measures for the policy positions of political actors are essential to testing foundational questions about political representation, partisan polarization, and electoral competition. Thousands of candidates run for the U.S. Congress each election season, yet we lack systematic information on the issue-level positions of the vast majority of these electoral contenders. To that end, we pair an original collection of campaign platforms with tools for machine learning to produce novel issue-level measures of congressional candidates' policy positioning. Our measures span multiple salient issue areas (e.g., guns, immigration, and abortion) for candidates who ran for the U.S. House of Representatives between 2018 and 2022. Through a series of validation tests, we demonstrate that our Candidate Positioning Indexes (CPIs) reliably capture latent issue-specific positions from text. We go on to demonstrate that unidimensional "ideal points" for candidate policy positioning mask important variation in issue polarization at the party level and multidimensionality in left-right positioning at the candidate level.

Keywords: Text scaling; Primary elections; Measurement; U.S. Congress; Multidimensionality; Polarization

---

*Authors are listed alphabetically, both contributed equally.

We are thankful for comments and feedback from Daniel Butler, Jaclyn Kaslovsky, Isaac Mehlhaff, Lucia Motolinia, Benjamin Noble, Santiago Olivella, Joesph Ornstein, and Julia Park, as well as participants in the 2023 Society for Political Methodology Annual Meeting and 2024 Vision in Methodology Conference. Elizabeth Borowiak, Ryan Casey, Erik Johansen, and Audrey Sachleben provided excellent research assistance. Unite America provided funding for elements of this project.

†Assistant Professor, Department of Political Science, University of Iowa, 359 Schaeffer Hall, Iowa City, IA 52242, colin-case@uiowa.edu.

‡Notre Dame du Lac Assistant Professor, Department of Political Science, University of Notre Dame, 2057 Jenkins Nanovic Halls, Notre Dame, IN 46556, rachel.porter@nd.edu.

Introduction

Measures for the policy positions of political actors are essential to theoretical and empirical political science. In spatial representations, observed behaviors of political actors determine their locations along a dimension of interest relative to other political actors. Resulting estimates are said to reflect actors' policy positions[1] and are used to explore foundational theories in the discipline about representation, polarization, and electoral competition. An ever-growing literature leverages methodological innovations and novel data sources to estimate the spatial policy positions for numerous political entities, including voters (e.g., Tausanovitch and Warshaw 2013), interest groups (e.g., Bonica 2013), political parties (e.g., Laver et al. 2003; Slapin and Proksch 2008), legislatures (e.g., Shor and McCarty 2011), and legislators (e.g., Ansolabehere et al. 2001; Bateman et al. 2017; Rheault and Cochrane 2020).

Despite advancements in data and methodology, accurate and reliable measures for the issue-level policy positioning of U.S. congressional candidates remain elusive. Ideally, data for scaling positions should be sourced directly from campaigns, cover a wide range of issues, and include a large sample of candidates (Druckman et al. 2009). Such comprehensive data is not readily available for the thousands of congressional candidates who run each election cycle. Commonly used measures, instead, rely on perception-based data to infer politicians' positions (e.g., Stone and Simas 2010; Barbera 2015; Bonica 2014), which can yield inconsistent conclusions in substantive research (Tausanovitch and Warshaw 2017). Available sources for data that come directly from campaigns systematically exclude primary election candidates (e.g., Ansolabehere et al. 2001; Montagnes and Rogowski 2015); these candidates are a critical population of study as primaries have become the main venue for competition in modern congressional races.

Another concern with extant measures for candidate policy positioning relates to dimensionality. Existing measures summarize candidates' positions in unidimensional space, collapsing pol-

[1]This terminology follows work by Laver et al. (2003) and Benoit et al. (2009). Like these authors, we do not claim to measure true unobservable preferences but seek to scale observed behaviors reliably. Others have termed such estimates as ideology (e.g., Shor and McCarty 2011; Bonica 2014), policy representations (Broockman 2016), strategic positions (Case 2023), or political orientations (Tausanovitch and Warshaw 2017).

icy stances across multiple issues into an "ideal point" estimate—however, this scaling approach masks variation in candidates' left-right positions across different issues. Elise Stefanik (R-NY) illustrates this variation in her campaign positioning. During her first congressional campaign, Stefanik championed moderate policies, including an all-of-the-above approach to energy production and expanded access to higher education. She simultaneously took a hard-right stance on gun ownership, opposing any restrictions on the right to bear arms. Stefanik's policy positions also evolved during her time in office, particularly on immigration. Initially supportive of family reunification and expanding visa access, she later campaigned on hardline immigration policies, emphasizing border security and punitive measures. If candidates systematically adopt diverse left-right positions across issues like Stefanik, then unidimensional positioning measures will say little about candidates' stances on the issues themselves (Broockman 2016; Ahler and Broockman 2018). In this case, issue-level measures of policy positioning become essential for answering fundamental questions about representation, polarization, and electoral choice. For example, does partisan polarization among elites vary across issues (Jochim and Jones 2012; Moskowitz et al. 2024)? Do candidates tailor their issue-specific policy positions to local district considerations (e.g., Ansolabehere et al. 2001; Burden 2004)?

We produce novel estimates for congressional candidates' issue-specific policy positions that overcome the limitations of extant measures. Specifically, we scale the left-right positions of candidates who ran for the U.S. House of Representatives between 2018 and 2022 across six issue areas: abortion, education, energy, guns, healthcare, and immigration. To characterize candidates' issue-specific positions, we employ an original collection of campaign platforms scraped from candidates' campaign websites (n = 4,501 or 75% of all major-party, ballot-eligible candidates). To extract and scale latent policy positions from these text data, we construct a text-to-measure pipeline that pairs tools for supervised machine learning with a word embedding approach. In brief, we use an ensemble of classification models to identify campaign platform paragraphs that pertain to our issue areas of interest. Next, we locally estimate embeddings for words in our corpus (i.e., word embeddings) and candidates' issue-specific text (i.e., candidate-issue embeddings).

These candidate-issue embeddings capture the semantic meaning of all text from a given candidate in a specific year on a particular issue (e.g., an embedding for Elise Stefanik's position on abortion in 2022). To compute our positioning estimates, we compare the similarity of our candidate-issue embeddings to word embeddings indicative of that issue's left-most and right-most position. Greater similarity to right-most (left-most) word embeddings indicates a more extreme right (left) position. We refer to our resulting estimates as Candidate Positioning Indexes (CPIs).

We adhere to best practices when constructing our text-to-measure pipeline, justifying our measurement choices and validating our scaling procedure (see Park and Montgomery 2023). Specifically, we show that our ensemble text classifier performs well at extracting issue-relevant content for out-of-sample platform assessments. We also illustrate the conditions under which CPIs are (or are not) robust to alternative modeling specifications. To demonstrate the semantic validity of our final estimates, we compare CPIs to human judgments of left-right policy positioning in campaign platform text. As a final test of convergent validity, we show that candidates who take far-left or far-right stances on issues more often receive PAC contributions from special interest groups that champion corresponding policy positions.

We conclude our paper by demonstrating the utility of our CPI estimates. Specifically, we uncover variability and multidimensionality in the issue-specific policy positioning of congressional candidates that have gone largely undocumented in extant work. At the party level, we show that the extent of issue polarization between Democrats and Republicans varies across issue areas and across time. At the candidate level, we uncover variation in issue-specific policy positioning—politicians are not always consistent in their left-right position-taking. Controlling for a candidate's own partisanship and the partisan leanings of her district, we find that a candidate's issue-specific policy positions track with district opinion on that specific policy area. We uncover no such relationship when employing a unidimensional measure of candidate policy positioning.

Beyond our descriptive findings, the text data from candidates' campaign platforms and issue-specific measures for candidate policy positioning made available through this project constitute a major contribution to the study of congressional politics. These resources open numerous avenues for future research into ongoing debates about politicians' convergence on voter preferences (Fowler and Hall 2016; Ryan and Ehlinger 2023; Anderson et al. 2024), the extent of ideological choice among candidates (Montagnes and Rogowski 2015; Moskowitz et al. 2024), the strength of political parties (Aldrich 2011), and potential avenues for policy compromise in today's polarized Congress (Levendusky 2023). Moreover, our method could be applied to other kinds of policy dynamics among congressional candidates (e.g., hawk versus dove on foreign policy, bipartisan versus partisan in rhetorical strategy) and extended to other political actors (e.g., Supreme Court justices, bureaucrats, and presidents).

The Importance of Issue-Level Positioning Estimates

Models of policy positioning seek to reliably capture the spatial placement of political actors with a low-dimensional representation. Contemporary models of U.S. politics generally express policy positions as unidimensional—that is, as a single point-estimate along a left-right axis (e.g., Ansolabehere et al. 2001; Bonica 2014; Bateman et al. 2017; Rheault and Cochrane 2020).[2] This approach assumes that policy positioning on various issues collapses onto a single continuum. Some studies find this to be the case (e.g., Poole and Rosenthal 2011), lending credence to the idea that "ideology scores" summarize policy positioning and aptly simplify complex political phenomena. However, other work doubts that unidimensional measures accurately reflect a political actor's positions on individual issues. In particular, Broockman (2016) argues that a unidimensional measure for policy positions captures an individual's degree of positional consistency across policy domains (e.g., an individual holds liberal views on two-thirds of issues) but says little about views on issues themselves. For instance, in a unidimensional framework, someone who is ideologically mixed (i.e., an individual who holds liberal and conservative preferences on policy) can appear as moderate as an individual who has consistently centrist views.

Research on candidates' position-taking motivations suggests these actors may adopt varying left-right stances across different issues. A host of factors influence a candidate's positioning on

[2] Early research on dimensionality and complexity in policy space expressed issues in n-dimensions (e.g., Clausen and Cheney 1970; Shepsle and Weingast 1994).

policy, such as her district's two-party competitiveness (Burden 2004), voters' preferences (Erikson and Wright Jr. 1980), and competitors' characteristics (Case 2023; Porter et al. 2024), as well as her own partisanship (Erikson and Wright Jr. 1989; Ansolabehere et al. 2001) and valence qualities (Stone and Simas 2010). Even as party brands have grown increasingly meaningful in U.S. politics, congressional candidates still have an incentive to distinguish themselves on a policy dimension; standing out from co-partisans is integral to winning increasingly competitive primaries. Indeed, some work shows that congressional candidates take divergent positions on especially salient issues (e.g., Porter 2022; Malzahn and Hall 2024). Other work similarly finds evidence for this kind of across-issue positional heterogeneity among political elites (e.g., Crespin and Rohde 2010; Roberts et al. 2016) and subsets of voters (e.g., Fowler et al. 2023).

If, as this literature suggests, congressional candidates engage in issue-specific moderation or extremism in their policy positioning, unidimensional scores will be inadequate for testing foundational questions about representation. For instance, existing research shows that certain constituents, particularly primary voters, are highly concerned with politicians' stances on specific issues, and these policy preferences can significantly influence vote choice (Ryan and Ehlinger 2023; Henderson et al. 2022). Notably, Ahler and Broockman (2018) demonstrate that citizens prefer candidates who align with them on issues as opposed to candidates whose overall "ideology" is closer to their own. This suggests that evaluating theories of proximity voting requires estimates for both voters *and* candidates' positions on issues themselves. Estimates of issue-specific positions are available for voters (Warshaw and Rodden 2012) and some legislators (Fowler and Hall 2016; Moskowitz et al. 2024). However, to our knowledge, there are no readily available estimates for the issue-specific policy positions of congressional candidates. Thus, it is unknown to what extent candidates maintain consistent left-right positions on policy and whether constituent preferences or other electoral factors can explain issue-specific moderation or extremity.

In addition to obscuring politicians' issue-specific policy positions, unidimensional measures mask the extent of intra-party conflict and inter-party polarization. To address growing partisan polarization in congressional politics, identifying when elite issue polarization attenuates and, importantly, the policy domains where moderation is likely to occur is a necessary first step. As Levendusky (2023) writes, "It may not be possible to bridge the gap on every issue, but there are cases where compromise and coming together is possible, and finding those pathways is worthwhile" (p. 154). Pathways for policy compromise may be found along issues where the parties are less deeply divided and along issues where parties are more internally heterogeneous. Unidimensional positioning measures are unsuitable for identifying such issue areas. Aldrich et al. (2014) raise this point when examining the dimensionality of scaled roll-call votes in the U.S. Congress, concluding that "a one-dimensional dominant result may reflect party 'teamsmanship'...we can tell only that parties are divided from one another, but not if they are divided on one issue, many issues, or...none at all" (p. 438). Through a series of simulations, the authors find that defining a multidimensional policy space is necessary to identify between and within-party policy cleavages. Jochim and Jones (2013) find similar evidence of across-issue variation in intra-party unity and inter-party polarization on policy when scaling roll call votes.

To that end, electoral rhetoric may provide greater nuance into issue-by-issue polarization in modern congressional politics than examinations of legislative roll-call votes. Approaches for scaling policy positions from voting behavior rely on issue-specific votes to generate estimates for legislators' issue-specific policy preferences. However, many salient issues do not receive an individual vote in a given session of Congress, either because these issues are left off the agenda or because the relevant legislative text is included in omnibus bills that cover multiple topics (Clinton 2012; Fowler and Hall 2016). Today's era of strong party leadership exasperates this selection bias as roll-call votes highlighting intra-party divisions are a rarity, leading to misconceptions about the pervasiveness of party homogeneity in congressional politics (Lee 2018; Duck-Mayr and Montgomery 2023). This is of concern for scholars because the measurement of inter- and intra-party polarization is central to evaluating theories of party strength, factionalism, and bipartisan coalition building. For all these reasons, other venues for policy positioning, such as elections, may offer an alternative perspective on the distribution of preferences between and within parties. However, to achieve this, novel data and innovative methodological strategies are needed to facilitate the measurement of politicians' issue-specific policy positions in congressional elections.

Our Measurement Approach

In this section, we outline our data and methodological approach for measuring issue-specific policy positions among candidates for U.S. Congress. We rely on an original collection of campaign platforms drawn from congressional candidates' campaign websites to produce our measures for policy positioning. We collect this text for all available candidates who ran for the U.S. House of Representatives in 2018, 2020, and 2022. We produce Candidate Positioning Indexes (CPIs) for six issue areas: abortion, education, energy, guns, healthcare, and immigration. These issues were identified by the Pew Research Center as top policy priorities among the general public and "very important" to vote choice among registered voters in election years of interest.[3]

Translating our corpus of campaign platforms into estimates for policy positioning in each issue area is a multi-stage process. Our text-to-measure pipeline proceeds as follows. First, we identify all paragraphs of text from each candidate's campaign platform that pertain to our six issue areas (e.g., all text about abortion from Elise Stefanik's campaign platform in 2022). Second, we estimate embeddings locally for every word in the vocabulary of candidates' campaign platforms. We simultaneously estimate embeddings for all issue-relevant text in each candidate's campaign platform (e.g., an embedding for Elise Stefanik's abortion-related platform text in 2022). Third, we measure the relative similarity of our candidate-issue embeddings to averaged word embeddings for dictionaries of terms indicative of the left-most and right-most policy position for a given issue area (e.g., pro-life versus pro-choice). These resulting similarity measures constitute our Candidate Positioning Indexes. In subsequent sections, we validate these estimates and demonstrate the utility of CPIs through a series of substantive applications.

Data: Campaign Websites

Per Druckman et al. (2009), an ideal data source on policy positioning in campaigns will be "unmediated (i.e., directly from the campaign), complete (i.e., covering a full range of rhetor-

[3]See Appendix A1 for greater detail on issue area selection.

ical strategies), and representative of the population of campaigns" (2009, p. 345). However, many sources for data on candidates' policy positions fail to meet these criteria. Some work on campaign policy positioning relies on surveys of congressional candidates, such as the National Political Awareness Test (NPAT) from Project Vote Smart (e.g., Ansolabehere et al. 2001; Montagnes and Rogowski 2015; Moskowitz et al. 2024). These surveys provide unmediated information on candidates' positions for a comprehensive set of issue areas but suffer from low response rates. For example, Montagnes and Rogowski (2015) report NPAT coverage for about 27% of non-incumbent, general election contenders. In pursuit of greater candidate coverage, other work relies on perception-based data to impute candidates' policy positions, such as expert opinion surveys (e.g., Stone and Simas 2010), follower networks (e.g., Barbera 2015), and fundraising patterns (Bonica 2014). Resulting measures extend to primary election contenders and may be suitable in some applications for scaling overall candidate positioning (see Tausanovitch and Warshaw 2017) but are unlikely to provide reliable estimates for candidates' positions on individual issue areas.

Digitization has broadened the accessibility of data suitable for measuring political actors' positions and preferences, providing new opportunities to reliably scale the spatial positions for a comprehensive set of congressional candidates. Text, in particular, has proven to be a promising data source as various latent concepts are embedded in and can be extracted from text (Laver et al. 2003; Slapin and Proksch 2008; Rheault and Cochrane 2020; Case 2023). For instance, recent work employs social media text to estimate candidate ideology (Bailey 2024; Hassell et al. 2023; Green et al. 2024). However, employing social media data has become more complicated and costly with recent roll-backs to academic researcher access (e.g., the elimination of Twitter's academic API and the shutdown of Meta's CrowdTangle platform).

In our measurement, we leverage text data drawn from congressional candidates' campaign websites to scale positioning on policy. Campaign websites are a source of data that uniquely fits the criteria proposed by Druckman et al. (2009). First, campaign websites are ubiquitous in modern campaigns. Porter et al. (2024) demonstrate that nearly all viable candidates today have a campaign website. Second, campaign websites express candidates' policy positions in their own words. Candidates and their staff spend substantial time crafting their website messaging because these sites serve as an informational "hub" in campaigns. Indeed, over a dozen states include links to campaign websites on official listings of ballot-eligible candidates—none provide links to social media.[4] Third, candidates take clear policy positions on their campaign websites across a host of issue areas. Existing work that compares the stances on a candidate's campaign website to her positions taken in other venues (i.e., speeches, debates, and advertisements) finds consistency in position-taking behavior (Sulkin et al. 2007).[5]

To produce our original collection of campaign platforms, we first compile a comprehensive list of all candidates running for the U.S. House in a given year and locate the URL for each candidate's campaign website, if available. Between 2018 and 2022, 91% of candidates had a campaign website. We characterize a candidate's campaign platform as the text presented on the "Issues" page on her campaign website. We next task teams of research assistants with extracting platform text from websites via manual downloading. All website text we employ was collected less than two weeks before a candidate's primary election to maximize data coverage and consistency. We document our data collection strategy in greater detail in Appendix A2.[6] Through this collection effort, we identify 75% of major-party, ballot-eligible candidates who ran in primary elections from 2018 to 2022 as having a campaign platform (4,501 of 6,006 total candidates).[7]

[4]As of 2024, these states include Alaska, Arizona, Delaware, Georgia, Hawaii, Illinois, Maryland, Minnesota, Montana, New Hampshire, Oregon, Vermont, and Virginia.

[5]More recently, work by Green et al. (2024) finds considerable variation in the partisan dimension of politicians' rhetoric across five different position taking venues. Importantly, this work focuses only on incumbents and does not examine campaign websites.

[6]We are not the first to collect text data from websites (e.g., Xenos and Foot 2005; Dolan 2005; Sulkin et al. 2007; Druckman et al. 2009; Bailey 2024; Meisels 2024; Pennec et al. 2024). However, our data collection effort is distinct in important ways. First, our population of interest constitutes *all* major-party, ballot-eligible candidates who ran in primaries or general elections. Other collections focus on collecting text for only a subset of all candidates. Second, we collected text in real-time, rather than relying on Internet archives via the Way Back Machine or Library of Congress, which can have idiosyncratic coverage. Third, we engage in thorough text cleaning and processing to allow for text aggregations at both the level of platform (full document) and platform point (sub-section).

[7]The proportion of congressional candidates present in our data surpasses the coverage of other major collections of electoral position-taking data, like the Wesleyan Media Project, which includes television advertisements for roughly 30% of all candidates in each election cycle, and Bonica's (2023) DIME database, which provides estimates for the ideological positioning of about 65% of candidates who ran between 2018 and 2022.

Method: Semantic Projection

Stage 1: Isolating Issue-Specific Text

When producing CPIs for our six issues of interest, we rely only on text that pertains to a specific issue area. Restricting the breadth of text used in our scaling procedure helps to limit noise in the measurement of candidates' left-right positions on policy (Grimmer and Stewart 2013; Egerod and Klemmensen 2020).[8] To isolate issue-specific text, we first splice campaign platforms into individual paragraphs.[9] Our corpus includes 136,368 natural paragraphs.[10] Next, we randomly sample 6% of these paragraphs (8,306 total) and label them for the presence/absence of content related to our six issue areas.[11] This categorization is not mutually exclusive; candidates occasionally discuss multiple issue areas within the same paragraph. Full coding instructions for labeling paragraphs can be found in Appendix A3.2.

We use labeled paragraphs to train a series of supervised machine-learning models that predict whether a given paragraph discusses a given issue area. We train five separate classification models for each of our six issues of interest and employ their predictions to train an ensemble stacking classifier. This approach allows us to leverage a diversity of modeling approaches to improve predictive accuracy in identifying issue-related paragraphs. Appendix A3.3 provides greater detail on model training and classifier validation. Across all issue areas, our ensemble classifier achieves an F1 score of at least 0.82 when classifying a held-out set of paragraphs from platforms.

[8] To substantiate this point further, we produce CPIs with a larger aggregation of text (i.e., full platform points rather than individual paragraphs). In Appendix A3.1, we demonstrate text inclusion has downstream consequences our estimates for candidate policy positioning.

[9] We focus our classification task on paragraphs rather than all the text nested under a specific issue header because candidates often discuss multiple issue areas in a single "platform point" (e.g., a candidate who discusses abortion and healthcare under "Women's Issues"). The choice of aggregation at the natural paragraph level follows best practices from Daubler et al. (2012) that emphasize the importance of exogenous unitization of text. Other extant work isolating message-specific content also aggregates text at the paragraph level (e.g., Grimmer and Stewart 2013).

[10] We define paragraphs as instances of line breaks. Paragraphs are sometimes exceedingly short (i.e., less than fifteen words). We aggregate short texts into the most proximate multi-sentence paragraph.

[11] Three expert readers labeled paragraphs for policy content; two coders labeled 3,000 paragraphs, and one coder labeled 2,306. Of the 8,306 paragraphs, 308 (4%) were labeled by all three coders to assess inter-coder reliability. For this reliability set, we achieve 88% agreement across all three readers.

Stage 2: Estimating Word Embedding Model

After isolating relevant paragraphs from candidates' campaign platforms, we employ a word embedding model to uncover meaning in these position-taking statements. Word embedding models use a neural network architecture to predict word(s) in a document given the word(s) that occur in close context to that word. Resulting embeddings are vector representations of semantic relationships between words in a dense, continuous, high-dimensional space. Using word embeddings to measure the ideological orientation of text has several advantages. First, unlike frequency-based scaling (e.g., Benoit et al. 2009; Slapin and Proksch 2008), an embedding-based approach relies on word co-occurrences to capture and represent relationships between words. Words appearing frequently in a similar context are assumed to have similar meanings; the word embedding training process captures this similar meaning by producing embeddings positioned nearby in space. This is useful for our purposes because, in campaign platforms, various words may be used interchangeably to make a similar semantic argument (e.g., "building a barrier on the southern border" versus "building a wall on the southern border"). Second, because word embeddings express word meaning in high-dimensional space, they better reflect the multidimensional use of words. In a policy context, this is important because certain words may take on different ideological bents based on context (e.g., "school choice" vs. "pro choice").

More specifically, our estimation approach relies on a Doc2Vec word embedding model. This approach is similar to a traditional word embedding model but allows for the incorporation of covariates in the training process (see Le and Mikolov 2014 for original specification); notably, the inclusion of covariates *produces covariate-level embeddings.* In our specification, each paragraph in our corpus is assigned a covariate that corresponds to the candidate author (e.g., Elise Stefanik), issue area (e.g., abortion), and election year (e.g., 2022).[12] Resulting covariate embeddings—which we refer to as candidate-issue embeddings[13]—are numerical representations of the rhetor-

[12] Recall that we classify some paragraphs (4% of all paragraphs) as discussing more than one issue (e.g., a paragraph discussing abortion and healthcare). In these cases, we include duplicate paragraphs in model training; each observation has a separate identifier (e.g., Elise Stefanik Abortion 2022 and Elise Stefanik Healthcare 2022).

[13] A more accurate descriptor would be "candidate-issue-year" embedding, as our approach produces unique embeddings for candidates who discuss the same issue across multiple elections. We use "candidate-issue" for brevity.

ical meaning for *all issue-specific paragraphs* from a given candidate in a given year (e.g., all paragraphs about abortion from Elise Stefanik in 2022). More simply, candidate-issue embeddings provide a semantic summary of a candidate's rhetoric on a specific issue of interest.

Our Doc2Vec model works in two parts: the first is the same as a traditional skip-gram Word2vec architecture (Mikolov et al. 2013). In this estimation, a word, w_k, is sampled at each iteration, where k is the word's position in a given document. A window of length Δ is extracted twice, once before and after w_k. The words in the window surrounding w_k are our outcomes of interest. These outcomes can be written more completely as $w_\Delta = (w_{k-\Delta},...,w_{k-1},w_{k+1},...,w_{k+\Delta})$. The model input is an indicator vector, x_k, for the target word, w_k, which selects the corresponding word embedding, β_k. This embedding predicts each word, m, in w_Δ. Figure 1 presents a graphical depiction of this word embedding model architecture.

The second part of our model uses the same architecture as above: a word, w_k, is sampled at each iteration, and a window of length Δ is extracted before and after. Once again, words falling in this window, w_Δ, represent our outcomes of interest. Instead of using an indicator vector to index the word embedding for target word w_k, the indicator vector, $z_{i,j,t}$, is used to index the candidate-issue embedding, $\zeta_{i,j,t}$, for candidate i on policy area j in election year t. This embedding predicts each word, m, in w_Δ.[14] Figure 2 provides a graphical depiction of this candidate-issue embedding model architecture.

In our implementation, we begin with embeddings pre-trained on the Google News corpus to set initial model weights for word embeddings; initial weights for candidate-issue embeddings are random. Our Doc2Vec model alternates between the first and second steps described above during training to tune the parameters within embeddings. We fit our model using parameter recommendations from Rodriguez and Spirling (2022), including a window of six and an embedding dimension of 300. We employ default hyper-parameters from the original Doc2Vec model. When

[14]This process is intuitively similar to taking the average of the word embeddings for candidate i on policy area j in year t. However, the Doc2Vec estimation places less weight on high-frequency words (e.g., the, is, in) that often possess little rhetorical meaning. For this reason, previous work finds Doc2Vec outperforms averaging word embeddings (Lau and Baldwin 2016; Grimmer et al. 2022) In Appendix A3.5, we see similar improvements when using Doc2Vec over averaging word embeddings on a series of validation tasks.

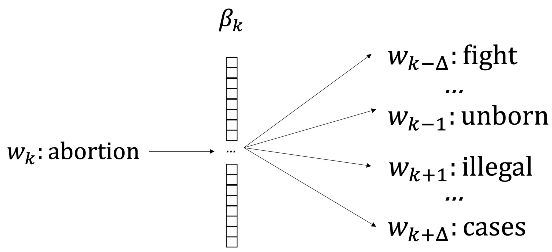


Figure 1: Word Embedding Architecture

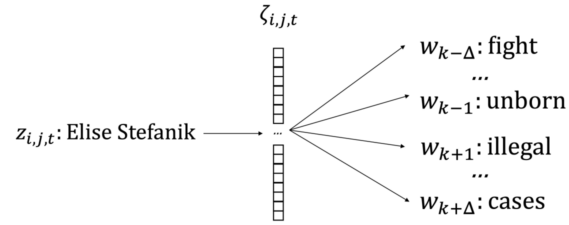


Figure 2: Candidate-Issue Embedding Architecture

training our model, we include only those paragraphs identified by our classifier as discussing issue areas of interest. Before model estimation, we take several standard text pre-processing steps laid out by Rodriguez and Spirling (2022).[15] These include removing words that do not appear at least ten times in our full corpus and converting text to lowercase.

Our trained model produces two sets of outputs. First, our model produces 11,577 word embeddings (β_k)—an embedding for every word in our corpus. Second, our model produces 18,630 covariate embeddings ($\zeta_{i,j,t}$)—an embedding for every candidate-issue-year combination in our data. If our classifier identifies no issue-relevant paragraphs for a given candidate in a given year, then we do not generate an embedding for that candidate-issue-year combination. Importantly, our modeling approach places covariate and word embeddings in the same dimensional space—we leverage this feature of model architecture to produce our CPI measures in the following section.

Stage 3: Scaling Candidate Policy Positioning

In this final stage, we scale candidates' policy positioning on a left-right dimension that defines modern-day discourse for our six issue areas of interest. Issues are multidimensional, and our measures do not capture positioning on all these dimensions; instead, we focus on a dominant cleavage. It is important to note that the flexible scaling procedure we introduce here can be exported to any number of policy cleavages. In Table 1, the text in italics denotes the left-right

[15]These pre-processing steps minimally alter the underlying meaning of text (Denny and Spirling 2018). However, in several important instances, these steps *do* affect the underlying meaning of words in our corpus. For this reason, we modify text in several instances to preserve underlying meaning (e.g., converting instances of "2nd amendment" to "second amendment"). The full list of our steps in text prep-processing is available in Appendix A3.4.

Table 1: Issue Cleavages & Positional Endpoint Term Dictionaries

Issue Area	Left-Most & Right-Most Term Dictionaries
Abortion	*Pro-Choice* (L): reproductive, justice, prochoice, codifying, autonomy, personal... *Pro-Life* (R): birth, unborn, conception, heartbeat, prolife, ban, born, outlaw...
Education	*Federal* (L): public, invest, free, tuition, universal, prek, equality, salaries, teachers... *Local* (R): parents, local, control, decentralize, choice, competition, charter, private...
Energy	*Renewables* (L): renewable, net, zero, incentives, credits, subsidize, develop... *Fossil Fuels* (R): oil, coal, deregulate, repeal, free, market, domestic, reserves...
Guns	*Restrictions* (L): mandatory, background, ban, national, registry, assault, automatic... *Access* (R): infringed, inherent, unrestricted, repeal, reciprocity, concealed, carry...
Healthcare	*Publicly Funded* (L): universal, medicareforall, singlepayer, human, right... *Privatized* (R): free, market, open, competition, choice, consumers, deregulate...
Immigration	*Inclusive* (L): undocumented, daca, pathway, dreamers, dignity, abolish, ice... *Exclusive* (R): enforce, law, build, wall, verify, sanctuary, chain, birthright...

cleavage that is the focus of our measurement for each issue area.

We employ the word and covariate embeddings estimated from the model described above to produce our issue-specific measures. We compare the relative closeness of each candidate's covariate embedding, $\zeta_{i,j,t}$, to the left-most and right-most positions for a given issue cleavage. We define these positional endpoints using dictionaries of terms and their associated word embeddings. We produce twelve dictionaries—a left-most and right-most term list for each of our six issue areas. Using dictionaries to capture key quantities of interest is standard in research employing word embeddings (e.g., Kitagawa and Shen-Bayh 2024; Garg et al. 2018; Garten et al. 2018). Table 1 outlines an abbreviated list of terms in each dictionary; Appendix Table A11 shows a full list of terms. The selection of these terms is based on the close reading of policy platforms from far-left and far-right advocacy groups. We discuss our term selection strategy in greater detail in Appendix A3.6 and demonstrate that our estimated CPIs are robust to alternative term selections.

The scaling procedure we employ can be thought of as semantic projection. It is well established that applying vector algebra to word embeddings produces a semantic axis of meaning (e.g., Bolukbasi et al. 2016; Garten et al. 2018; Kozlowski et al. 2019; Grand et al. 2022). In our scaling procedure, we subtract the word embeddings for terms in our left and right issue dictionaries to produce a single embedding representing the left-right dimension of an issue cleavage. We project a candidate's issue-specific text, represented by their candidate-issue embedding, onto this axis. The resulting estimate constitutes our Candidate Positioning Index.[16] As CPI increases (decreases), a candidate's text is closer in meaning to the right-most (left-most) position for that issue cleavage.

The full scaling procedure for constructing CPIs is as follows. First, for issue j, we average the embeddings for the set of words, R_j and L_j, in our right and left issue dictionaries, respectively. Resulting averages produce two embeddings—ρ_j, and λ_j—indicative of the semantic meaning of the right-most and left-most positions for the cleavage defining j. Second, we subtract λ_j from ρ_j to create an axis of left-right positioning, α_j. Finally, for each candidate i in year t with an estimated covariate embedding for issue j, we calculate the cosine similarity between $\zeta_{i,j,t}$ and α_j. This estimation procedure for calculating our Candidate Position Indexes is written as:

$$
\begin{aligned}
R_j &= \{w_1, w_2 ... w_n\} \\
L_j &= \{w_1, w_2 ... w_n\} \\
\rho_j &= \overline{\beta_{k \in R_j}} \\
\lambda_j &= \overline{\beta_{k \in L_j}} \\
\alpha_j &= \rho_j - \lambda_j \\
CPI_{i,j,t} &= cos(\alpha_j, \zeta_{i,j,t})
\end{aligned}
$$

Measurement Validation

We compare human judgments of policy platform text to our Candidate Positioning Indexes as a benchmark for validity. Our validation framework seeks to uncover whether CPIs match human readers' placement of policy-specific campaign platform text on a left-right dimension. This validation exercise is a meaningful test for the semantic validity of our estimates (Grimmer

[16] We standardize CPI with a mean of 0 for ease of interpretation.

and Stewart 2013; Lowe and Benoit 2013). In this task, three expert readers assess the same random selection of documents (900 documents per reader). Expert readers include undergraduate and graduate students. Documents constitute paragraphs aggregated by candidate-issue-year;[17] these texts are identical to those we employ to estimate candidate-issue embeddings. Readers are blind to any candidate or party-identifying information beyond what was available in text documents. Readers are provided detailed instructions outlining each issue cleavage of interest, as well as broad guidelines about scoring and policy-specific examples of extremity in left-right positioning. These instructions are included for review in Appendix Section A4.1.

Referencing these instructions and relying on their own judgments, readers score documents on a five-point scale from very left (–2) to very right (+2).[18] We limit the potential range of values for scoring to whole numbers. We considered several alternative specifications for validation, such as crowd-sourced pairwise comparisons (Benoit et al. 2016; Carlson and Montgomery 2017) or LLM classification (Ornstein et al. 2024). After repeated testing, our task specification produced the highest scoring quality for our latent quantities of interest at the lowest cost.[19] Full documentation of this human judgment task and descriptive statistics on reader performance and scoring are provided in Appendix A4. Sampled texts are included in online Supplementary Materials.

Figure 3 presents a series of coefficient plots with results from this validation exercise. We estimate OLS models with human judgment as our dependent variable, measured as a document's average score across all three readers.[20] Facets of Figure 3 evaluate the association between CPI and human judgments both within and across party for the issue areas of abortion, guns, and immigration (validation for remaining issue areas is ongoing). Full outputs for models presented

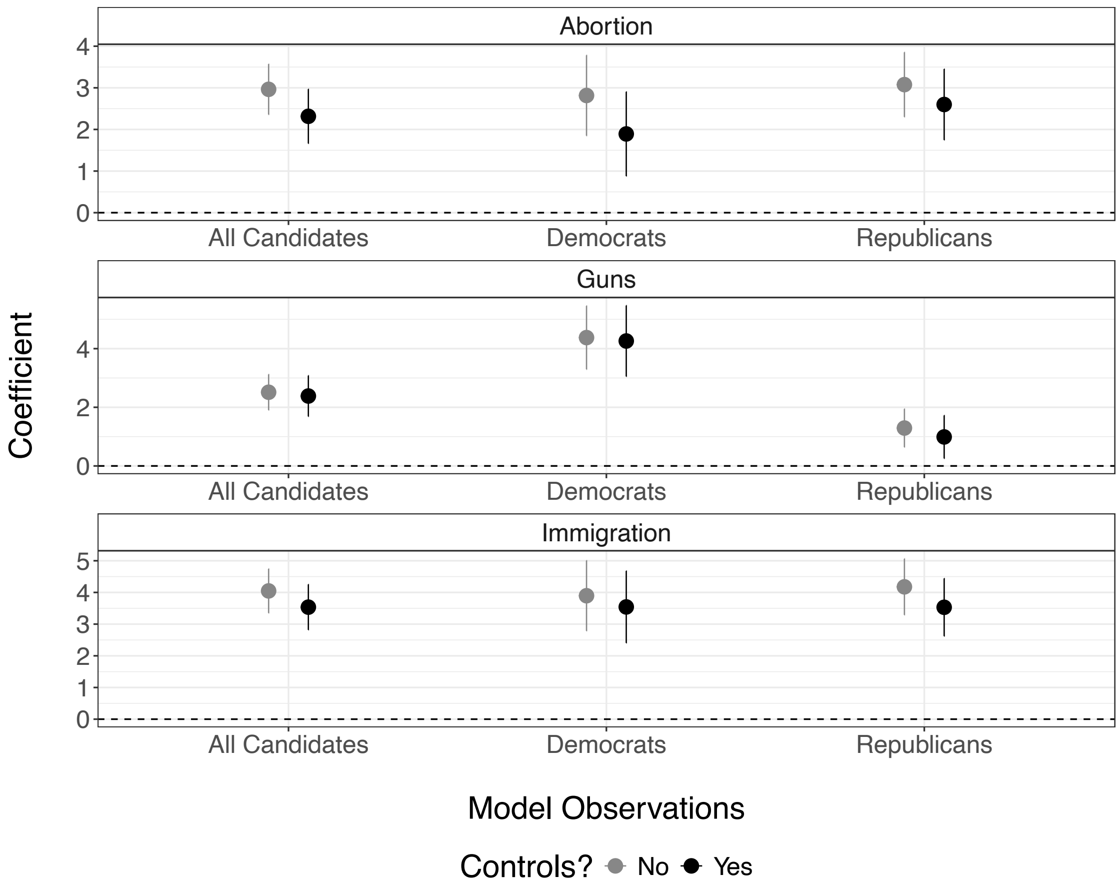
[17]For clarity, if candidate i had multiple paragraphs identified for a given year t about issue j, these paragraphs were collapsed into a single document for readers.

[18]Readers also flagged document content ambiguous in positioning or irrelevant to scaling our quantities of interest. For the purposes of this validation task, these scores were set to zero (i.e., centrist). Descriptive statistics on ambiguous position-taking and the presence of irrelevant documents are provided in Appendix Tables A15 and A16.

[19]Research finds that crowd-sourced task workers increasingly rely on LLMs to complete tasks (Veselovsky et al. 2023), which could produce a lower quality of scores for our purposes. Our prompting of Open AI's GPT-3.5 produced unstable estimates for positioning extremity in a pairwise comparison framework.

[20]In Appendix Table A17, we estimate an identical set of models using an ordered logistic regression. Here, the dependent variable is the modal score assigned by the three readers. The lead reader's classification is assigned in instances where all three readers are divided; this occurred in less than 5% of assigned documents.

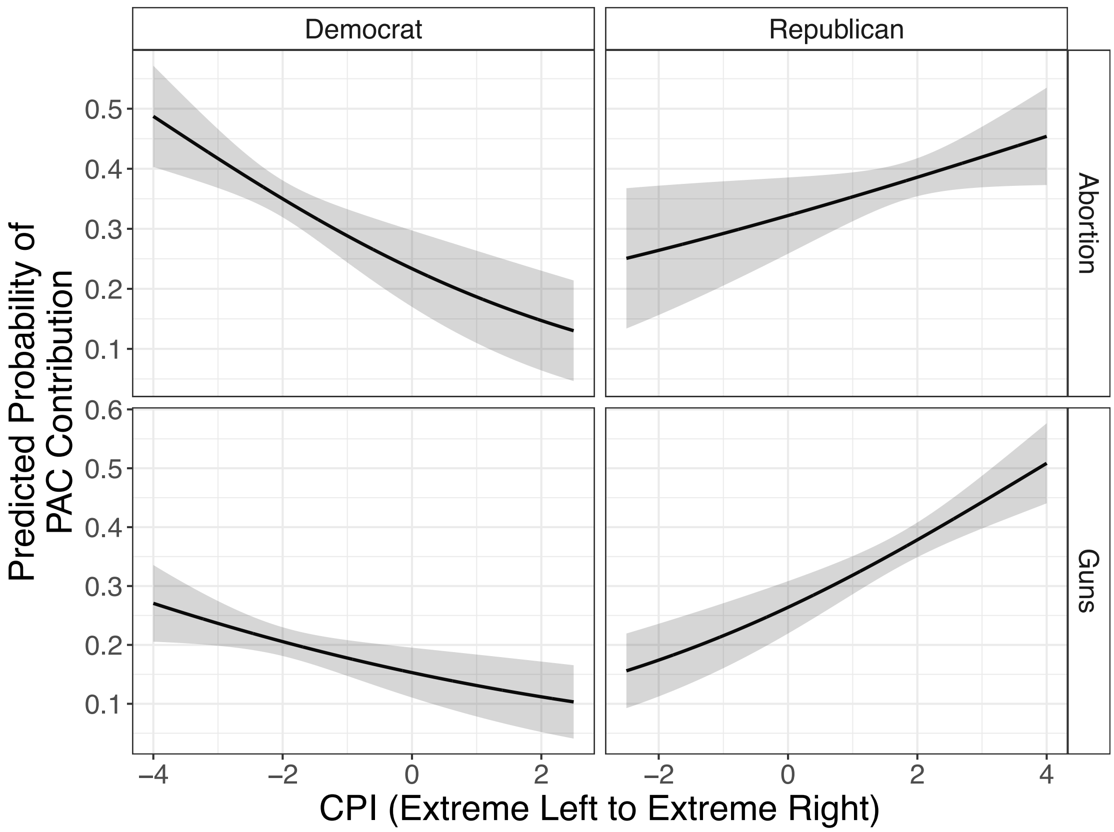
Figure 3: CPI Validation with Human Judgements



in Figure 3 are available in Appendix Table A19.

Turning first to bivariate regression in Figure 3, these results suggest that human readers and our scaling approach pick up the same latent qualities in text. To further demonstrate both the validity and utility of our estimates, we estimate a second series of regressions that control for unidimensional positioning, as estimated by Bonica (2014), as well as candidate party when applicable. Here, we assess whether CPI captures human judgments after accounting for a candidate's broader positioning. Across these control models, CPIs serve as a statistically significant predictor for human judgments of text. In Appendix Table A18, we replicate this analysis with an alternative measure for unidimensional policy positioning. We use Case's (2023) measure for unidimensional policy positioning, produced using the same campaign website data introduced in this paper. We find substantively identical results as those presented in Figure 3.

Figure 4: CPI Validation with Position-Aligned PAC Giving



As an additional test of convergent validity, we explore the relationship between Political Action Committee (PAC) giving and CPI extremity. Extensive research documents the giving behavior of special interest groups, finding that these entities contribute to candidates for access-oriented and ideological reasons (e.g., Fouirnaies and Hall 2014; Barber 2016; Meisels 2024). Politicians may even tailor their position-taking behavior to generate greater appeal to donors and groups (Baker 2016). For our purposes, we are agnostic to the direction of the causal arrow between candidate positioning and donor giving. The underpinning of our validation exercise is that candidates whose issue positions more closely align with the priorities of special interest groups should more often receive fundraising contributions from these groups (Li 2018; Bonica and Li 2021). If CPIs are reliable and accurate measures for policy-specific positioning in elections, we should expect an association between a candidate's extremity in left-right policy positioning and her likelihood of garnering fundraising from a position-aligned PAC.

Figure 4 presents predicted probabilities for PAC giving as a function of candidate positioning extremity. The dependent variables across logistic regressions are binary indicators for the presence/absence of a PAC donation from an abortion-related and gun-related special interest group. When examining pro-choice and pro-gun control PACs giving behavior, we constrain units of analysis to Democratic candidates; when examining pro-life and pro-gun rights PACs giving behavior, we constrain units of analysis to Republican candidates.[21] We identify PACs as aligned with the pro-choice/pro-life and gun control/rights positions using OpenSecrets data. We include fundraising data on all giving in support of a candidate (i.e., direction donations and independent expenditures). Models control for unidimensional positioning, as estimated by Bonica (2014).[22] Full model outputs are available in Appendix Tables A20 to A23. Across all four panels of Figure 4, we find a statistically significant association between PAC giving and the extremity of issue-specific candidate positioning. Recall, negative (positive) values indicate more extreme left (right) policy positioning. For our Democratic and Republican candidate models, the association between PAC giving and positional extremity is in the expected direction.

Descriptive Results

Having walked through our scaling procedure and validated our CPI estimates, we now use empirical applications to descriptively evaluate multidimensionality and variation in congressional candidates' policy positions. First, we investigate party-level issue polarization; we find substantial variation across issues in the extent to which Democratic and Republican candidates are polarized. Additionally, we demonstrate cross-temporal variation in issue polarization at the party level. We go on to assess the degree to which candidates take consistent left-right positions and find only modest pairwise correlations across issue areas. Finally, we evaluate whether candidates' policy positioning varies systematically with district conditions. Controlling for the partisan leanings of a candidate's district and her own political orientations, we find that a candidate's issue-specific

[21]Units include only candidates for whom we estimate CPI-Abortion or CPI-Guns; as discussed earlier, we do not generate issue-specific CPIs for candidates with no issue-specific text. We replicate this analysis with imputed scores in Appendix Tables A20 to A23. These results are substantively similar to those presented in Figure 4.

[22]Bivariate models produce substantively identical results and can be found in Appendix Tables A20 to A23.

policy positions frequently track with district opinion on closely related policies.

Party-Level Variation in Issue Polarization

Political elites in the United States have polarized over the last several decades. Polarization across the parties is documented in Congress (e.g., Poole and Rosenthal 1985, 2011), as well as in congressional elections (e.g., Bonica 2014; Case 2023). Less is known, however, about variation in party polarization at the issue level. Some existing research measures issue-level partisan polarization (e.g., Jochim and Jones 2013; Fowler and Hall 2016; Moskowitz et al. 2024). However, due to data and methodological constraints, extant work is limited in the scope of issues examined, types of politicians studied, and/or time periods considered. Our Candidate Positioning Indexes address these limitations, allowing us to explore the distribution of policy positions taken by incumbents and challengers in primaries across various salient political issues in recent elections.

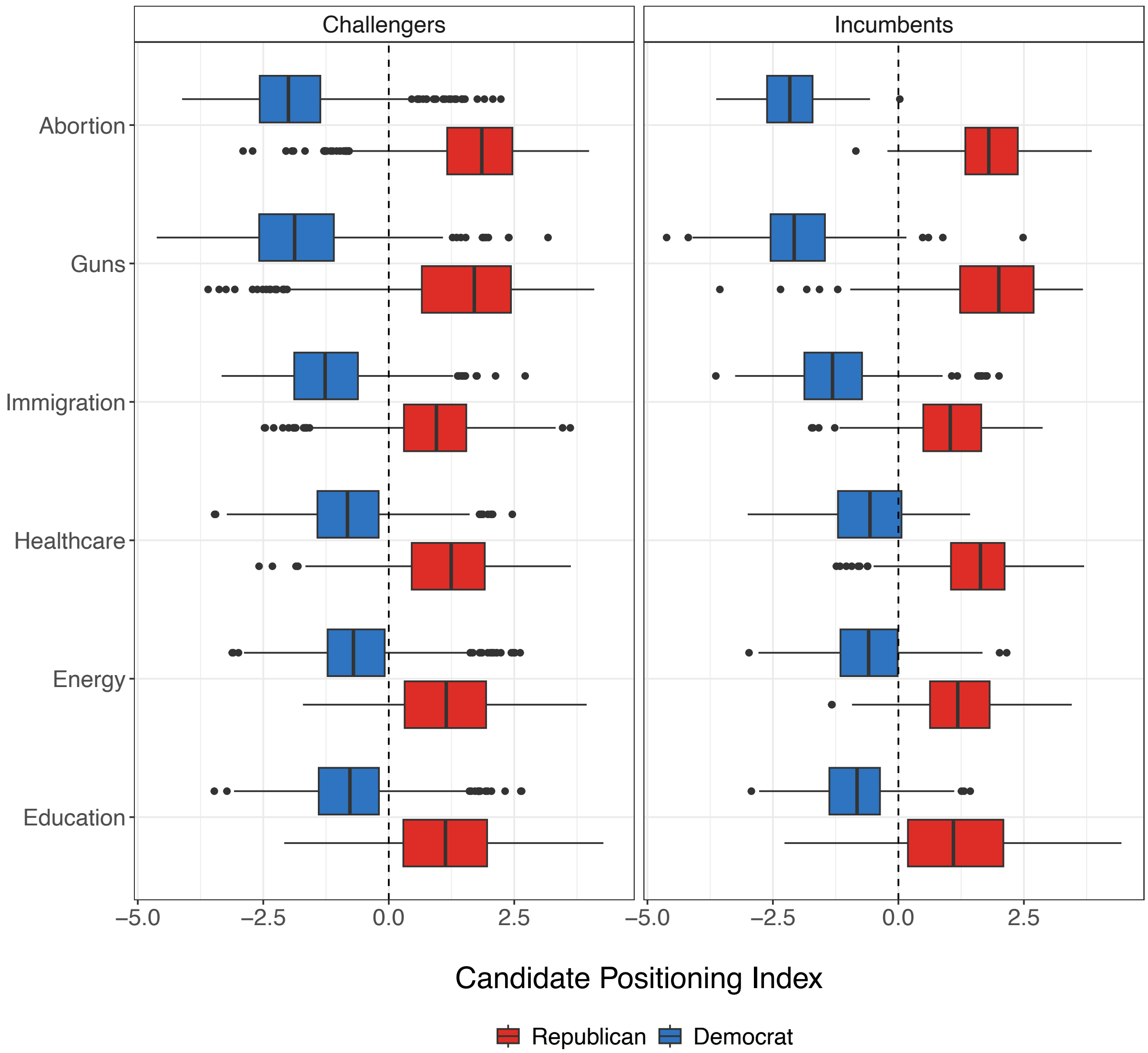
Figure 5 depicts CPI distributions by party, pooled across time and issue area. Policy plots are sorted from most polarized (top) to least polarized (bottom). The right facet of Figure 5 depicts distributions of CPI for incumbents running for reelection; the left facet depicts distributions for challengers (i.e., non-incumbents). There is striking issue-by-issue variation in polarization between the parties and positional heterogeneity within parties. On some issues (e.g., guns), we observe greater inter-party polarization and intra-party variance in positioning; for other issue areas (e.g., healthcare), there is lower inter-party polarization and intra-party variance in CPI estimates. Interestingly, we observe *little* difference in issue polarization when comparing the distributions of incumbents and challengers.

Figure 6 further explores intra-party variation in issue polarization by plotting party mean CPIs for several issue areas by year.[23] In Republican Party (Democratic Party) facets, large positive (negative) values indicate greater distance from a moderate position, centered at 0.[24] Per Figure 6, there is significant variation in the magnitude and direction of shifts in party mean CPIs across time. For instance, the party means for CPI-Energy became more extreme in recent elections for both

[23] A plot with all issue areas is available in Appendix Figure A9

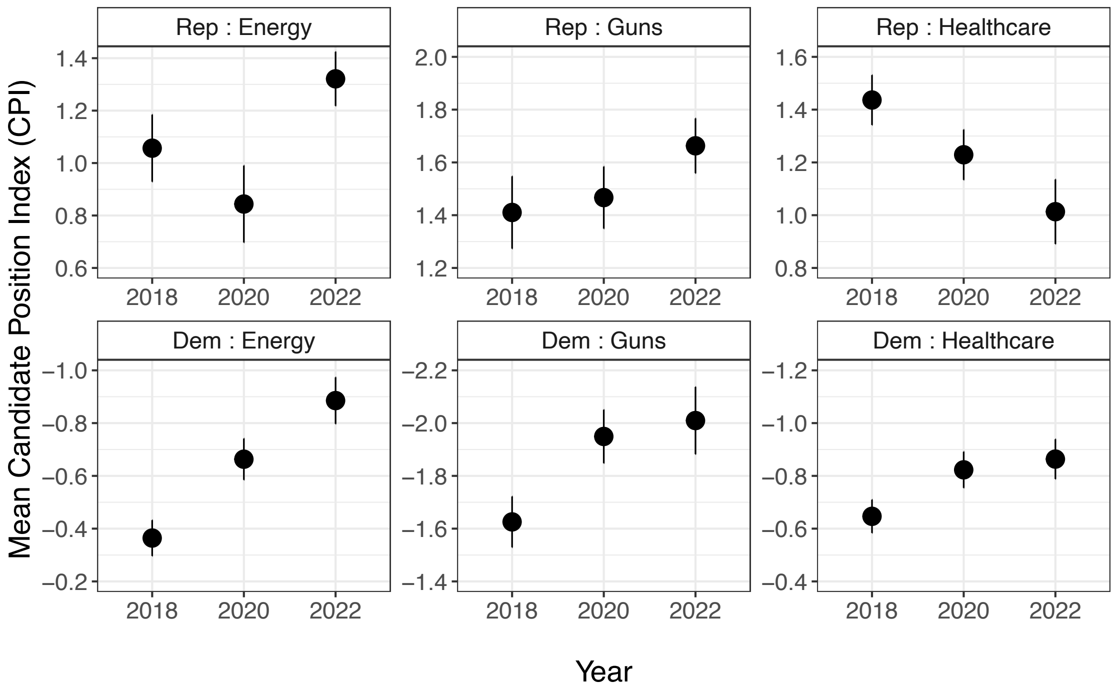
[24] The y-axis for Democrat facets in Figure 6 is inverted to parallel the directional extremity in Republican facets.

Figure 5: Box Plot of Party-Level CPI, By Issue and Candidate Type



the Democratic and Republican parties. These shifts are substantively noteworthy; the magnitude change is 60% of a standard deviation in the distribution of CPI-Energy for the Democratic Party and 55% of a standard deviation for the Republican Party. Notably, party shifts in issue polarization are not equivalent for all policy areas. For CPI-Guns, we observe between-party differences in the magnitude of extremity shifts; the increase in extremity for the Republican Party CPI-Gun mean (a 15% standard deviation increase) is less than half the magnitude of the Democratic Party shift (a 40% standard deviation increase). Still other issues see little change in extremism. For CPI-Healthcare, the Democratic Party's mean CPI shifted nominally between 2018 and 2022. On the

Figure 6: Candidate Positioning Index Party Mean, By Issue and Year



other hand, the Republican Party's mean CPI for this issue became *less extreme* over each election cycle, with a magnitude decrease of almost half of a standard deviation.

The primary takeaway from these analyses is that partisan polarization varies at the issue level and across time—importantly, this nuance in partisan polarization is obscured in unidimensional measures. Our findings align with a growing body of work suggesting that politicians are indeed less divided on issues than conventional wisdom suggests (e.g., Bateman et al. 2017; Curry and Lee 2020). We show that, even for policy areas with greater differences in partisan distributions, policy preferences still overlap. These findings also align with growing methodological concerns regarding the use of roll call votes to document partisan polarization; existing research suggests that selection bias in the roll call record (Ainsley et al. 2020), party discipline (Lee 2018; Duck-Mayr and Montgomery 2023), and inconsistency in the issue agenda (Moskowitz et al. 2024) may obscure potential pathways for compromise and consensus that exist among lawmakers. Our findings echo this sentiment, underscoring that text data provides a fruitful path forward for studying issue-level polarization.

Candidate-Level Variation in Policy Positioning

There is a long-standing and ongoing scholarly debate regarding the utility of unidimensional measures of policy positioning (e.g., Heckman and Snyder Jr. 1996; Poole and Rosenthal 2011; Aldrich et al. 2014; Broockman 2016; Ahler and Broockman 2018; Fowler et al. 2023). Some argue that policy preferences across multiple issues collapse well on a single left-right spectrum. Others find that policy positions are unconstrained and multidimensional; thus, unidimensional measures capture policy consistency but say little about views on issues themselves. This literature almost exclusively examines voters' policy preferences; far fewer studies assess whether politicians' positions are well-captured in unidimensional space.

Understanding the extent to which politicians' positions vary across issues also has normative importance. A central question to scholars of political representation and electoral politics concerns whether candidates advocate for policy preferences in a way that reflects district opinion (e.g., Ansolabehere et al. 2001; Burden 2004). This question today has taken on renewed importance as congressional elections have nationalized. Over the past several decades, both major parties have invested heavily in cultivating a unified party brand, carefully crafting their messaging to define and dramatize their differences (Lee 2016; Hopkins 2018). These endeavors to consolidate party messaging have produced their intended effect: America's two major political parties are today perceived to offer voters starkly polarized policy alternatives. In this context, it is often assumed that congressional candidates across the country run on the same party-driven platforms, offering voters uniform policy options that pay little regard to local political dynamics.

To that end, we investigate the extent of systematic positional variation across our Candidate Positioning Indexes. We observe only modest correlations between our issue-specific CPIs. Cross-issue correlations in CPIs are strong when examining all candidates, ranging from 0.67 to 0.84. However, these correlations diminish when examining positional consistency within each party, ranging from 0.20 to 0.38. To unpack this heterogeneity, we assess whether the observed variation in left-right positions of congressional candidates varies with a candidate's local political context. We are particularly interested in determining whether the issue-specific positions adopted by

Table 2: Cooperative Election Policy-Specific Survey Questions

Issue Area	Survey Question: Do you support or oppose...?
Abortion	*Pro-Choice* (L): Always allowing a woman to obtain an abortion as a matter of choice *Pro-Life* (R): Permitting abortions ONLY in extreme cases (e.g., rape, incest)
Guns	*Restrictions* (L): Banning assault rifles *Access* (R): Making it easier to obtain a concealed-carry gun permit
Healthcare	*Publicly Funded* (L): Providing Medicare (Medicare-for-all) for all Americans *Privatized* (R): Repealing the entire Affordable Care Act
Immigration	*Inclusive* (L): Grant legal status to all undocumented immigrants (2020/22 only) *Exclusive* (R): eliminating the visa lottery and decreasing legal immigration by 50%

congressional candidates track with district-level opinion on related policy areas (e.g., candidate pro-choice positioning as a function of district-level support for pro-choice policy).

We fit a series of regressions to explore the relationship between a candidate's district context and her policy positioning. The dependent variables in our main analyses are issue-specific CPIs. Our primary independent variables are congressional district-level measures for policy opinions on topics related to our estimated CPIs. Table 2 outlines the left-right policy questions from the 2018, 2020, and 2022 Cooperative Elections Surveys for which we develop district-level public opinion estimates. We pair these survey questions with a multilevel regression and postratification (MRP) model to produce estimates of each congressional district's policy opinions. A host of studies demonstrate that MRP outperforms disaggregated survey responses in producing accurate public opinion estimates for small geographic units (e.g., Lax and Phillips 2009; Warshaw and Rodden 2012).[25] For our estimations, we adopt Warshaw and Rodden's (2012) modeling specification; the estimates we produce are robust to alternative model specifications and estimation strategies.[26]

[25]In brief, the first stage of MRP involves modeling survey responses as a function of demographic and geographic predictors, allowing for the partial pooling of information across congressional districts. The second stage involves weighting model-estimated demographic-geographic respondent types using census-level data on their prevalence in each district's population. For a full review see Lax and Phillips (2009) and Warshaw and Rodden (2012).

[26]We specify a multilevel regression poststratified with the US Census's American Community Survey 5-year data. We do not use the one-year surveys because no survey was fielded in 2020 due to the COVID-19 pandemic. Following Warshaw and Rodden (2012), we poststratify our estimates using the gender:education:race joint distributions; replicating our estimation with age:education:race joint distributions produces nearly identical estimates. We considered estimating our models with a partial or full-Bayesian implementation in Stan. We elect not to use this implementation because it exponentially increased model estimation time with little change in output.

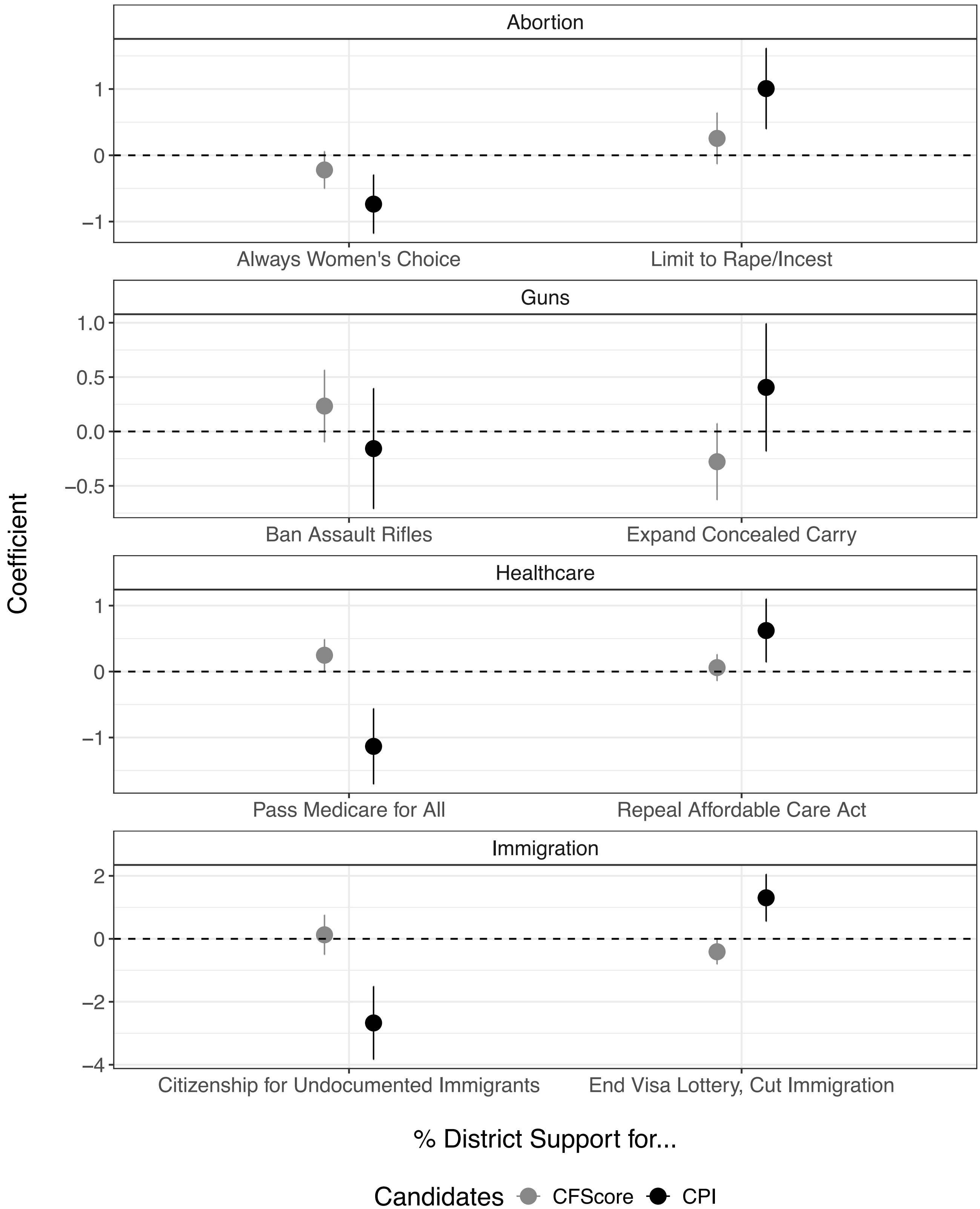
Our models control for other standard candidate-level predictors of position-taking, including candidate partisanship, district partisanship, incumbency, past elected experience, and gender. We also specify our models with year-state fixed effects.[27] Importantly, we estimate a second series of models using Bonica's 2014 unidimensional CF scores as the dependent variable. Our goal here is to evaluate whether our CPIs uncover a relationship between local district conditions and candidate policy positioning that is unobserved with a unidimensional positioning measure. Full model outputs for all analyses are available in Appendix Tables A24-A27.

Figure 7 plots model coefficients for the effect of district policy opinion on candidate CPIs across various issue areas and model specifications. The x-axis denotes the policy content of survey questions; complete question wording is available in Table 2. The y-axis denotes the directionality and magnitude of these regression coefficients for district opinion. Positive coefficients indicate that as a district becomes more right-leaning in its policy stance (e.g., pro-life, pro-gun rights), a candidate takes a more right-leaning stance in their policy positioning. Negative coefficients indicate that as a district becomes more left-leaning in its policy stance (e.g., pro-choice, pro-gun safety), a candidate takes a more left-leaning stance in their policy positioning. Coefficient estimates are displayed with 95% confidence intervals.

For models predicting CPIs (black coefficients), a consistent association exists between district policy-specific opinion and candidate issue-specific positioning. As a district becomes more favorable towards a right (left) leaning policy, candidate positioning on that issue tracks in a more extreme right (left) leaning direction. For models predicting CFscores (grey coefficients), there is *no* statistically significant or substantively noteworthy association between district policy-specific opinion and candidate "ideological" positioning. The results presented here underscore two key findings. First, candidates' policy positioning is not well-captured on a single left-right dimension. Second, candidates' positional variation tracks with district-level factors—a relationship that is unobserved with unidimensional positioning measures.

[27] We do not observe sufficient candidate-level variation in district opinion to estimate a two-way fixed effect design.

Figure 7: Relationship Between District Policy Opinion and Candidate Issue Positioning



Discussion & Future Research

This article introduced a novel text-to-measure pipeline for extracting and scaling latent policy positions within texts. We produced estimates for the issue-specific policy positions for three-quarters of all major-party, ballot-eligible candidates who ran for the US House between 2018 and 2022. Our measurement extends to six salient policy areas, including immigration, abortion, and gun control. Importantly, our measurement strategy is flexible insofar as it can be exported to any number of other policy applications in the context of US politics and beyond. To produce our positioning estimates, we employ an original compilation of campaign platforms drawn from congressional candidates' campaign websites. A lack of comprehensive and reliable data on the policy positions of congressional candidates has impeded research on fundamental theories of political science regarding representation, issue polarization, and electoral competition. The text data used in this project is a contribution in and of itself and will be made open-source to advance the study of representation in American legislative politics.

We aim to provide full documentation of our carefully constructed text-to-measure pipeline so to provide objective evidence about the validity of our scaling procedure and final estimates. We validate and justify our measurement decisions at each stage of estimation. For our final CPI estimates, we illustrate the semantic and convergent validity of our positioning measures through multiple validation exercises to show that our measures (1) reliably uncover latent policy positions within texts, and (2) map onto other salient political phenomena. We hope this article and its supplementary materials serve as resource for scholars seeking to scale latent qualities in text.

Across multiple applications, we demonstrate the utility of our Candidate Positioning Indexes (CPIs). We show that CPIs uncover previously undocumented multidimensionality and variation in electoral policy positioning at both the party and candidate levels. We find that issue polarization in elections is variable; parties are less deeply divided on certain policy areas (e.g., healthcare and energy) and more internally divided on others (e.g., education and guns). At the candidate level, we find that cross-issue correlations in candidates' left-right positions are modest at best. Importantly, we find that this variation in policy positioning tracks with district-level policy opinion.

References

Ahler, Douglas J., and David E. Broockman. 2018. "The Delegate Paradox: Why Polarized Politicians Can Represent Citizens Best." *Journal of Politics* 80(4).

Ainsley, Caitlin, Clifford J. Carrubba, Brian F. Crisp, Betul Demirkaya, Matthew J. Gabel, and Dino Hadzic. 2020. "Roll-call vote selection: implications for the study of legislative politics." *American Political Science Review* 114(3): 691–706.

Aldrich, John H. 2011. *Why Parties? A Second Look*. University of Chicago Press.

Aldrich, John H., Jacob M. Montgomery, and David B. Sparks. 2014. "Polarization and ideology: Partisan sources of low dimensionality in scaled roll call analyse." *Politial Analysis* 22(4): 435–456.

Anderson, Sarah E., Daniel M. Butler, and Laurel Harbridge-Yong. 2024. "The Primary Premium: Why Legislators are Incentivized to Side with Primary Voters over General Election Voters.".

Ansolabehere, Stephen, James M. Snyder Jr., and Charles Stewart III. 2001. "Candidate Positioning in U.S. House Elections." *American Journal of Political Science* 45(1).

Bailey, Michael A. 2024. "Measuring Candidate Ideology in Congressional Elections, 2020-2022.".

Baker, Anne E. 2016. "Do Interest Group Endorsements Cue Individual Contributions to House Candidates?" *American Politics Research* 44(2).

Barber, Michael. 2016. "Donation Motivations: Testing Theories of Access and Ideology." *Political Research Quarterly* 69(1): 148–159.

Barbera, Pablo. 2015. "Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data." *American Political Science Review* 23(1): 76–91.

Bateman, David A., Joshua D. Clinton, and John S. Lapinski. 2017. "A House Divided? Roll Calls, Polarization, and PolicyDifferences in the U.S. House, 1877–2011." *American Journal of Political Science* 61(3): 698–714.

Benoit, Kenneth, Drew Conway, Benjamin E.Lauderdale, Michael Laver, and Slava Mikhaylov. 2016. "Crowd-sourced Text Analysis: Reproducible and Agile Production of Political Data." *American Political Science Review* 110(2): 278–295.

Benoit, Kenneth, Michael Laver, and Slava Mikhaylov. 2009. "Treating Words as Data with Error: Uncertaintyin Text Statements of Policy Positions." *American Political Science Review* 53(2): 495–513.

Bolukbasi, Tolga, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. "Man is to computer programmer as woman is to homemaker? debiasing word embeddings." *Advances in neural information processing systems* 29.

Bonica, Adam. 2013. “Ideology and interests in the political marketplace.” *American Journal of Political Science* 2(57): 294–311.

Bonica, Adam. 2014. “Mapping the Ideological Marketplace.” *American Journal of Political Science* 58(2): 367–386.

Bonica, Adam, and Zhao Li. 2021. “Campaign Agendas and Issue Group Strategy in Congressional Primaries.”.

Broockman, David E. 2016. “Approaches to studying policy representation.” *Legislative Studies Quarterly* 41(1): 181–215.

Burden, Barry C. 2004. “Candidate Positioning in US Congressional Elections.” *British Journal of Political Science* 34: 211–227.

Carlson, David, and Jacob M. Montgomery. 2017. “A Pairwise Comparison Framework for Fast, Flexible, and Reliable Human Coding of Political Texts.” *American Political Science Review* 111(17): 835–843.

Case, Colin R. 2023. “Measuring Strategic Positioning in Congressional Elections.”.

Clausen, Aage R., and Richard B. Cheney. 1970. “A comparative analysis of Senate-House voting on economic and welfare policy, 1953–1964.” *American Political Science Review* 64(1): 138–152.

Clinton, Joshua D. 2012. “Using roll call estimates to test models of politics.” *Annual Review of Political Science* 15(1): 79–99.

Crespin, Michael H., and David W. Rohde. 2010. “Dimensions, issues, and bills: Appropriations voting on the House floor.” *Journal of Politics* 72(4): 976–989.

Curry, James M., and Frances E. Lee. 2020. *The Limits of Party: Congress and Lawmaking in a Polarized Era*. University of Chicago Press.

Daubler, Thomas, Kenneth Benoit, Slava Mikhaylov, and Michael Laver. 2012. “Natural sentences as valid units for coded political texts.” *British Journal of Political Science* 42(4): 937–951.

Denny, Matthew J., and Arthur Spirling. 2018. “Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It.” *Political Analysis* 26(2): 168–189.

Dolan, Kathleen. 2005. “Do women candidates play to gender stereotypes? Do men candidates play to women? Candidate sex and issues priorities on campaign websites.” *Political Research Quarterly* 58(1): 31–44.

Druckman, James N., Martin J. Kifer, and Michael Parkin. 2009. “Campaign Communications in U.S. Congressional Elections.” *American Political Science Review* 103(3): 343–366.

Duck-Mayr, JBrandon, and Jacob Montgomery. 2023. “Ends against the middle: Measuring latent traits when opposites respond the same way for antithetical reasons.” *Political Analysis* 31(4): 606–625.

Egerod, Benjamin CK, and Robert Klemmensen. 2020. "Scaling political positions from text: Assumptions, Methods and Pitfalls." In *he SAGE Handbook of Research Methods in Political Science and International Relations*. SAGE Publications pp. 498–521.

Erikson, Robert S., and Gerald C. Wright Jr. 1980. "Policy representation of constituency interests." *Political Behavior* 2: 91–106.

Erikson, Robert S., and Gerald C. Wright Jr. 1989. Voters, Candidates, and Issues in Congressional Election. In *Congress Reconsidered, 4th edition*. pp. 37–90.

Fouirnaies, Alexander, and Andrew B. Hall. 2014. "The Financial Incumbency Advantage: Causes and Consequences." *Journal of Politics* 76(3): 711–724.

Fowler, Anthony, and Andrew Hall. 2016. "The Elusive Quest for Convergence." *Quarterly Journal of Political Science* 11(1): 131–149.

Fowler, Anthony, Seth J. Hill, Jeffrey B. Lewis, Chris Tausanovitch, Lynn Vavreck, and Christopher Warshaw. 2023. "Moderates." *American Political Science Review* 117(2): 643–660.

Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. "Word embeddings quantify 100 years of gender and ethnic stereotypes." *Proceedings of the National Academy of Sciences* 115(16): E3635–E3644.

Garten, Justin, Joe Hoover, Kate M Johnson, Reihane Boghrati, Carol Iskiwitch, and Morteza Dehghani. 2018. "Dictionaries and distributions: Combining expert knowledge and large scale textual data content analysis: Distributed dictionary representation." *Behavior research methods* 50: 344–361.

Grand, Gabriel, Idan Asher Blank, Francisco Pereira, and Evelina Fedorenko. 2022. "Semantic projection recovers rich human knowledge of multiple object features from word embeddings." *Nature Human Behavior* 6: 975–987.

Green, Jon, Kelsey Shoub, Rachel Blum, and Lindsay Cormack. 2024. "Cross-Platform Partisan Positioning in Congressional Speech." *Political Research Quarterly* .

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

Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. 2022. *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton University Press.

Hassell, Hans, Michael Heseltine, and Kevin Reuning. 2023. "Buying What Candidates are Selling: Campaign Activity and Voters' Perceptions of Candidate Ideology.".

Heckman, James J., and James M. Snyder Jr. 1996. "Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators." *National Bureau of Economic Research Working Paper* .

Henderson, John A., Geoffrey Sheagley, Stephen N. Goggin, Logan Dancey, and Alexander G. Theodoridis. 2022. "Primary divisions: How voters evaluate policy and group differences in intraparty contests." *Journal of Politics* 84(3): 1760–1776.

Hopkins, Daniel J. 2018. *The Increasingly United States: How and Why American Political Behavior Nationalized*. University of Chicago Press.

Jochim, Ashley E., and Bryan D. Jones. 2012. "Issue Politics in a Polarized Congress." *Political Research Quarterly* 66(2).

Jochim, Ashley E., and Bryan D. Jones. 2013. "Issue politics in a polarized congress." *Political Research Quarterly* 66(2): 352–369.

Kitagawa, Risa, and Fiona Shen-Bayh. 2024. "Measuring Political Narratives in African News Media:A Word Embeddings Approach." *Journal of Politics* .

Kozlowski, Austin C., Matt Taddy, and James A. Evans. 2019. "The Geometry of Culture: Analyzing Meaning through Word Embeddings." *American Sociological Review* 84(5): 905–949.

Lau, Jey Han, and Timothy Baldwin. 2016. "An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation." *1st Workshop on Representation Learning for NLP* .

Laver, Michael, Kenneth Benoit, and John Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 97(2): 311–331.

Lax, Jeffrey R., and Justin H. Phillips. 2009. "How Should We Estimate Public Opinion in the States?" *American Journal of Political Science* 53(1): 107–21.

Le, Quoc, and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *International conference on machine learning*. PMLR pp. 1188–1196.

Lee, Frances E. 2016. *Insecure Majorities: Congress and the Perpetual Campaign*. Chicago, IL: University of Chicago Press.

Lee, Frances E. 2018. "The 115th Congress and Questions of Party Unity in a Polarized Era." *Journal of Politics* 80(4): 1464–1473.

Levendusky, Matthew. 2023. *Our Common Bonds: Using What Americans Share to Help Bridge the Partisan Divide*. Chicago University Press.

Li, Zhao. 2018. "How Internal Constraints Shape Interest Group Activities: Evidence from Access-Seeking PACs." *American Political Science Review* 112(4): 792–808.

Lowe, Will, and Kenneth Benoit. 2013. "Validating Estimates of Latent Traits from Textual Data Using Human Judgment as a Benchmark." *British Journal of Political Science* 21: 298–313.

Malzahn, Janet, and Andrew B. Hall. 2024. "Election-Denying Republican Candidates Underperformed in the 2022 Midterms." Forthcoming, *American Political Science Review*.

Meisels, Mellissa. 2024. “Campaign Agendas and Issue Group Strategy in Congressional Primaries.”.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. “Efficient estimation of word representations in vector space.” *arXiv preprint arXiv:1301.3781* .

Montagnes, Brendan Pablo, and Jon C. Rogowski. 2015. “Testing Core Predictions of Spatial Models: Platform Moderation and Challenger Success.” *Political Science Research and Methods* 3(3): 619–640.

Moskowitz, Daniel J., Jon C. Rogowski, and James M. Snyder Jr. 2024. “Parsing Party Polarization in Congress.” *Quarterly Journal of Political Science* .

Ornstein, Joseph T, Elise N Blasingame, and Jake S Truscott. 2024. “How to train your stochastic parrot: Large language models for political texts.”.

Park, Ju Yeon, and Jacob M. Montgomery. 2023. “Towards a framework for creating trustworthy measures with supervised machine learning.”.

Pennec, Caroline Le, Rafael Di Tella, Randy Kotti, and Vincent Pons. 2024. “Keep your Enemies Closer: Strategic Candidate Adjustments in U.S. and French Elections.”.

Poole, Keith T., and Howard L. Rosenthal. 2011. *Ideology and congress*. Transaction Publishers.

Poole, Keith T., and Howard Rosenthal. 1985. “A Spatial Model for Legislative Roll Call Analysis.” *American Journal of Political Science* 29(2): 357–384.

Porter, Rachel. 2022. *Some Politics Are Still Local: Strategic Position Taking in Congress and Elections*. Phd thesis, University of North Carolina at Chapel Hill.

Porter, Rachel, Sarah A. Treul, and Maura McDonald. 2024. “Changing the Dialogue: Descriptive Candidacies and Position Taking in Campaigns for the US House of Representatives.” *Journal of Politics* .

Rheault, Ludovic, and Christopher Cochrane. 2020. “Word embeddings for the analysis of ideological placement in parliamentary corpora.” *Political Analysis* 28(1): 112–133.

Roberts, Jason M., Steven S. Smith, and Stephen R. Haptonstahl. 2016. “The Dimensionality of Congressional Voting Reconsidered.” *American Politics Research* 44(5): 794–815.

Rodriguez, Pedro L., and Arthur Spirling. 2022. “Word embeddings: What works, what doesn’t, and how to tell the difference for applied research.” *Journal of Politics* 84(1): 101–115.

Ryan, Timothy J., and J. Andrew Ehlinger. 2023. *Issue Publics: How Electoral Constituencies Hide in Plain Sight*. Cambridge University Press.

Shepsle, Kenneth, and Barry Weingast. 1994. “Institutional foundations of committee power.” *American Political Science Review* 81(1): 85–104.

Shor, Boris, and Nolan McCarty. 2011. "The ideological mapping of American legislatures." *American Political Science Review* 105(3): 530–551.

Slapin, Jonathan B., and Sven-Oliver Proksch. 2008. "A Scaling Model for Estimating Time-Series Party Positions from Texts." *American Journal of Political Science* 52(3): 705–722.

Stone, Walter J., and Elizabeth N. Simas. 2010. "Candidate Valence and Ideological Positions in U.S. House Elections." *American Journal of Political Science* 54(2): 371–388.

Sulkin, Tracy, Cortney M. Moriarty, and Veronica Hefner. 2007. "Congressional Candidates' Issue Agendas On- and Off-line." *The International Journal of Press/Politics* 12(2): 63–79.

Tausanovitch, Chris, and Christopher Warshaw. 2013. "Measuring Constituent Policy Preferences in Congress, State Legislatures, and Cities." *Journal of Politics* 75(2): 330–342.

Tausanovitch, Chris, and Christopher Warshaw. 2017. "Estimating Candidates' Political Orientation in a Polarized Congress." *Political Analysis* 25(2): 167–187.

Veselovsky, Veniamin, Manoel Horta Ribeiro, and Robert West. 2023. "Artificial Artificial Artificial Intelligence: Crowd Workers Widely Use Large Language Models for Text Production Tasks.".

Warshaw, Christopher, and Jontahan Rodden. 2012. "How Should We Measure District-Level Public Opinion on Individual Issues." *Journal of Politics* 74(1): 203–19.

Xenos, Michael A., and Kirsten A. Foot. 2005. "Politics As Usual, or Politics Unusual? Position Taking and Dialogue on Campaign Websites in the 2002 U.S. Elections." *Journal of Communication* 55(1): Pages 169–185.

Supplementary Materials for: Measuring Policy Positioning in U.S. Congressional Elections

Colin R. Case and Rachel Porter

Contents

A1 Issue Area Selection

To select the six issue areas that define our issue-specific measures for US congressional candidate policy positioning, we turn to a series of surveys from the Pew Research Center. Every election year, Pew asks registered voters about a series of issue areas and how important each issue is to that voters' vote choice in November. They also ask the general public about issues they view as a "top priority" for Congress. Below, we report the proportion of registered voters who cite these issues as "very important" to vote choice and the proportion of the general public who cite these issues as a "top priority" for Congress.

Election Year—2018
Public Priorities: https://www.pewresearch.org/politics/2018/01/25/economic-issues-decline-among-publics-policy-priorities/

- Abortion —
- Education (72%)
- Energy (Environment) (62%)
- Energy (Climate Change) (46%)
- Guns —
- Healthcare (Healthcare Cost) (68%)
- Healthcare (Medicare) (66%)
- Immigration (47%)

Election Year—2018
Voting Priorities: https://www.pewresearch.org/politics/2018/10/04/2018-midterm-voters-issues-and-political-values/

- Abortion (61%)
- Education —
- Energy —
- Guns —
- Healthcare (58%)
- Immigration (68%)

Election Year—2020
Voting Priorities: https://www.pewresearch.org/politics/2020/08/13/important-issues-in-the-2020-election/

- Abortion (40%)
- Education —
- Energy (Climate Change) (42%)
- Guns (55%)
- Healthcare (68%)

- Immigration (52%)

Election Year—2022
Public Priorities: https://www.pewresearch.org/politics/2022/02/16/publics-top-priority-for-2022-strengthening-the-nations-economy/

- Abortion —
- Education (58%)
- Energy (Climate Change) (42%)
- Guns —
- Healthcare (Healthcare Cost) (61%)
- Immigration (49%)
- Policing (Reducing Crime) (52%)

Election Year—2022
Voting Priorities: https://www.pewresearch.org/politics/2022/10/20/midterm-voting-intentions-are-divided-economic-gloom-persists/

- Abortion (56%)
- Education —
- Energy (61%)
- Guns (57%)
- Healthcare (63%)
- Immigration (54%)

A2 Data Collection

To collect text data from candidate campaign websites, we first identify the names of all major party candidates running in 2018, 2020, and 2022 using candidate filings with the Federal Election Commission (FEC) and state-level elections websites. Using this list of names, we seek to identify the campaign website URLs for all candidates in each election year by following links from online repositories like Politics1.com, visiting candidates' social media pages, and conducting simple Google searches. Candidates for whom we locate no campaign website are re-checked again in the days leading up to their primary to ensure we did not miss their site. Using this approach, we identify 5,478 of 6,006 major-party, ballot-eligible candidates who ran for the US House of Representatives between 2018 and 2022 as having a campaign website.

Once we identify the URL for a candidate's campaign website, we determine whether that candidate also had a policy platform on her website. We trace the process of collecting text from campaign platforms in Figure A8. For many candidates, this was a simple process; platform pages often had clear titles like "Where I Stand" or "My Positions." Of those 5,478 congressional candidates who had a campaign website, 82% also had a campaign platform—or 75% of all major-party, ballot-eligible candidates who ran between 2018 and 2022.

In Table A3, we explore predictors for missingness in our data. We estimate a linear probability model (left column) and logistic regression (right column) where the outcome variable is whether (1) or not (0) a given candidate had a campaign website with a policy platform. We include variables for incumbency status, previous office-holding experience, candidate party, total primary election fundraising (logged), primary election rules, whether the primary is contested, whether the seat was open, and election year. We find that running unopposed in a primary, being a Republican candidate, running in 2020, and being an office-holder are all statistically significant predictors for not having a campaign platform.

After locating a campaign website (Figure A8, panel 1) and campaign platform (Figure A8, panel 2), we locate the text for platform points. In nearly all cases, text presented on a candidate's campaign platform is organized as a series of platform points. We define a platform point as the body text nested under a descriptive text header. For some candidates, platform point text is included on a single page as a series of paragraphs broken up by headers denoting the issue area for that position (e.g., Abortion, Second Amendment Rights, or Agriculture). In the case shown in panel 3 of Figure A8, each platform point has its own dedicated sub-page.

We archive text from campaign platforms using a Qualtrics form. First, we record the unique identifier for the candidate (Figure A8, panel 4). Next, we collect meta-data for the candidate (Figure A8, panel 5), such as their race, gender, past elected experience, and fundraising identifier from the Federal Election Commission (FEC). We collect this information from a candidate's website; we rely on auxiliary resources (e.g., state legislative biographic profiles, newspaper articles, Wikipedia summaries, and Ballotpedia profiles) when needed. Finally, we archive text for each platform point (Figure A8, panel 6). We catalog each platform point as an individual "document" rather than saving all text in a single .txt file. In doing so, we assign the text a major-topic code aligning with policy areas from the Comparative Agendas Project, save the header for each platform point, and preserve platform point ordering.

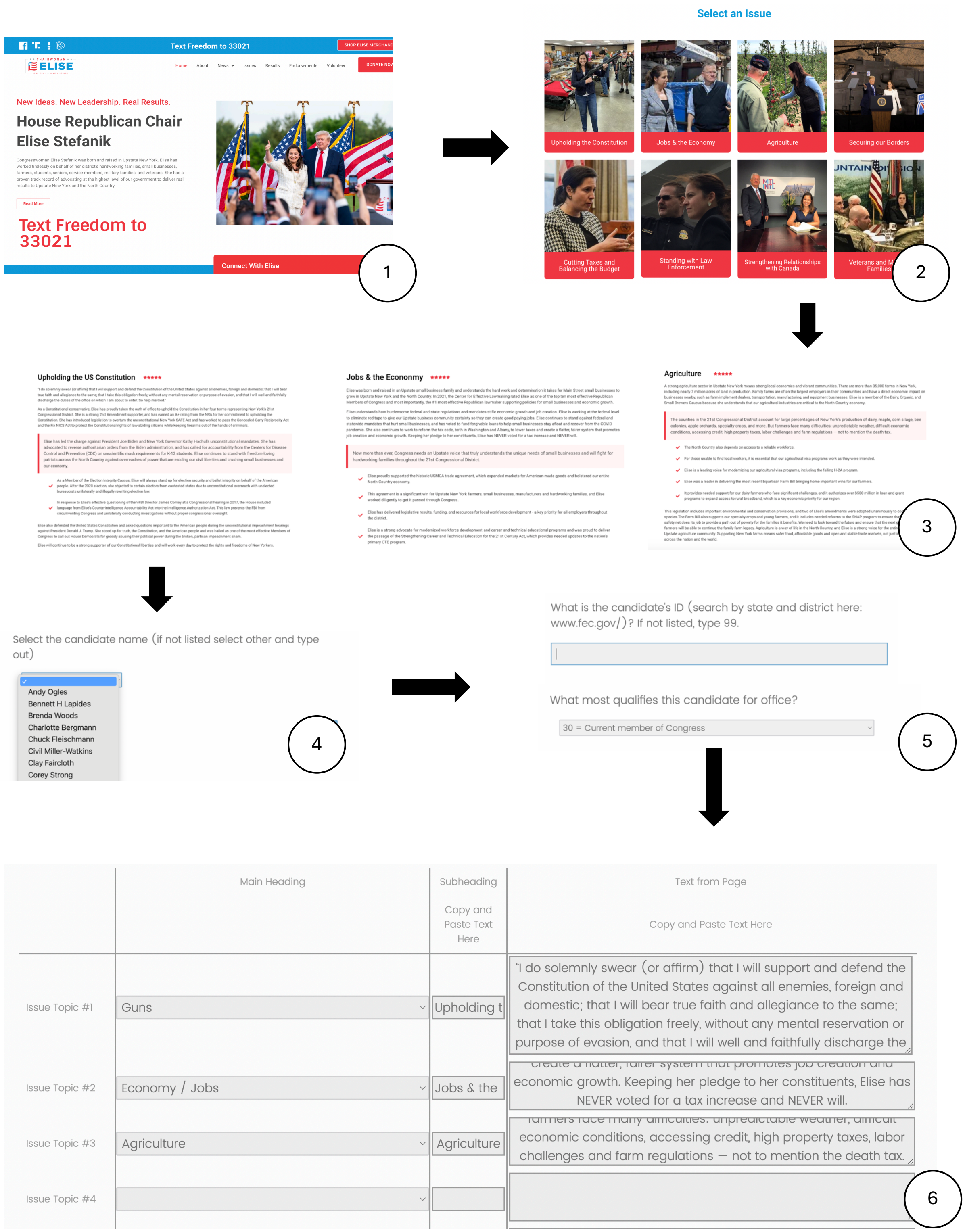
To ensure consistency in text collection, we scrape campaign websites during a two-week window before a candidate's primary election. We collect text before the primary because many candidates deactivate their websites immediately after losing their race. Internet archive repositories, such as the Way Back Machine or Library of Congress, catalog the campaign websites for some congressional candidates. However, many candidates' websites, particularly primary candidates' websites, are not archived. Further, text from "Issues" pages on archived sites is often not cataloged consistently. Web-scraping crawlers for the Way Back Machine only catalog for specific URLs at a fixed time. For example, if "https://www.ocasiocortez.com/" is scraped on a given day, this does not mean sub-pages on this site, such as "https://www.ocasiocortez.com/issues," are also scraped. Because of this, the homepage for a given candidate's campaign website may frequently be scraped by an archiving web crawler, but that candidate's campaign platform page is never cataloged or is only cataloged sparsely. For these reasons, we collect all text data in real time.

A small proportion of incumbents—less than 10% of incumbent members of the US who ran for reelection between 2018 and 2022—did not have a campaign website with a platform of policy positions. Most often, these campaign websites would contain only a single landing page that would function to solicit donations. For this minority of incumbents, we collect policy positions stated on their House.gov website. For individuals who do not wish to use scores created with these data, we denote all such observations with a dichotomous indicator in the version of our data available in this paper's replication materials.

Table A3: Predictors for Missingness in Campaign Platforms

	DV: Presence of Policy Platform	
	OLS	GLM
No Incumbent in Election	−0.017	−0.085
	(0.013)	(0.080)
Primary Type: Open	−0.019	−0.124
	(0.012)	(0.075)
Primary Type: Non-Partisan	−0.050*	−0.294*
	(0.017)	(0.102)
Unopposed Primary	−0.045*	−0.278*
	(0.016)	(0.095)
Republican Candidate	−0.027*	−0.173*
	(0.011)	(0.064)
Prior Office-Holder	−0.058*	−0.351*
	(0.017)	(0.108)
Current Incumbent MC	−0.143*	−0.871*
	(0.017)	(0.111)
Logged Pre-Primary Fundraising	0.029*	0.156*
	(0.001)	(0.007)
2020	−0.055*	−0.330*
	(0.013)	(0.078)
2022	−0.012	−0.079
	(0.013)	(0.080)
Constant	0.600*	0.520*
	(0.017)	(0.097)
Observations	6,006	6,006
R^2	0.109	
Log Likelihood		−3,062.394
Note:		*p<0.05

Figure A8: Steps for Campaign Website Text Collection



A3 Semantic Projection

A3.1 Testing Units of Aggregation

In the analyses below, we test the performance of CPIs generated using a different unit of text aggregation. Recall that the measurement framework presented in our main paper relies only on paragraphs of text identified by our issue classifier (see discussion in Section A3.3). We use this approach because some candidates discuss multiple issue areas under a single platform point heading (e.g., discussing abortion and health insurance under "Our Healthcare System"). Our goal is to isolate relevant paragraphs of text to limit noise in measurement, a step which follows existing work (e.g., Grimmer and Stewart 2013; Wilkerson and Casas 2017; Laver 2017).

Producing measures at different units of aggregation produces sufficiently different positioning estimates; we demonstrate this by comparing CPI performance to a positioning measure that includes non-relevant text. Our approach for producing this alternative measure follows the same estimation procedure used to generate CPIs, except we estimate positioning based on *all* text under a relevant platform point (e.g., all text under "Our Healthcare System"). We identify issue-relevant platform points as those that contain at least one issue-relevant paragraph identified by our classifier. Hereafter, we refer to this alternative measure as "CPI-Platform-Point." Correlations between CPI and CPI-Platform-Point are listed below; we report within and across-party correlations. Correlations range from relatively high (0.722) to low (0.146). These correlations indicate that text inclusion does impact the scaling of left-right positioning in policy platform text.

- Abortion: Cross-Party (0.563); Dems (0.473); Reps (0.722)
- Education: Cross-Party (0.673); Dems (0.278); Reps (0.721)
- Energy: Cross-Party (0.599); Dems (0.267); Reps (0.503)
- Guns: Cross-Party (0.773); Dems (0.450); Reps (0.672)
- Healthcare: Cross-Party (0.656); Dems (0.296); Reps (0.658)
- Immigration: Cross-Party (0.564); Dems (0.146); Reps (0.516)

The differences in measurement demonstrated above are attributable to text inclusion, given that we alter no other elements of measurement. There is sufficient reason to believe that these differences are specifically attributable to extraneous text inclusion. As we demonstrate in Section A3.3 of the Appendix, our text ensemble stacking classifier has recall and F1 scores of 0.80 or higher, indicating few cases of false negatives (i.e., text which are policy-relevant are not systematically being missed). From this analysis, we conclude that extraneous text inclusion has downstream consequences on measurement that may impact our scaling of policy positioning.

A3.2 Coding Instructions for Paragraph Classification

Identifying Paragraph Issue Content

This task involves reading paragraphs from congressional candidates' campaign platforms and judging whether the text pertains to issue areas of interest. If you believe the paragraph pertains to a given issue area, mark a "1" in that issue area column in the attached Excel spreadsheet. If you do not believe the paragraph pertains to a given issue area, leave that column in the attached Excel spreadsheet blank. Paragraph categorizations are not mutually exclusive; this is to say, paragraphs can be labeled for multiple issues of interest (e.g., one paragraph can pertain to both Abortion and Healthcare). Reference the text below to determine whether a given paragraph pertains to a given issue area:

Abortion—the ability to get an abortion

- Pro-life, pro-choice
- Not contraception, not prenatal healthcare, not adoptions
- "Reproductive healthcare" counts
- Planned Parenthood
    - We need to defund planned parenthood (yes)
    - We need to defund Planned Parenthood because it provides abortions (yes)
    - We need to defund Planned Parenthood because they sell fetal body parts (no)
    - We need to support planned parenthood (yes)
    - We need to support planned parenthood because it provides abortions (yes)
    - We need to support planned parenthood because it provides breast cancer screenings (no)

Education—ability to get an education, government involvement, what is taught in schools

- Increase teacher salaries, school choice, homeschooling, defunding the Department of Education, increasing the control of parents and local school boards, free college, college debt, reforming Common Core / No Child Left Behind

Energy—kinds of resources used to produce energy

- Fossil fuels (oil, natural gas, coal)
- Renewables (solar, wind, biomass, geothermal, hydroelectric, nuclear)
- All-of-the-above energy policy, energy independence
- Does not include discussions about just climate change, protecting the environment, conservation

Guns—ability to purchase weapons

- Gun access, gun rights (2nd amendment, limits on purchasing guns, no-fly list, assault weapons ban, red flag laws, concealed carry reciprocity)
- Discussions of school shootings, NRA alone not enough—must discuss gun access/restrictions

Healthcare—access to healthcare; cost of healthcare; treatments

- Any discussion of healthcare (Obamacare, insurance), rural healthcare access, inequalities in quality of healthcare, women's healthcare, Medicare
- Abortion included if the discussion refers to "abortion as healthcare; just pro-life, pro-choice does not count; may reference the healthcare services provided by Planned Parenthood
- Includes COVID-19; includes discussions of vaccines; does not include lockdowns, businesses closing

Immigration—the ability to enter the country, discussions of immigrants within the country

- Immigration reform; immigrants coming over the border; border security; immigrants seeking asylum; DACA; Dreamers; services to be provided to immigrants; pathways to citizenship
- This can include broader discussions of the treatment of immigrants, immigrant experience, sanctuary cities, deportation, ICE, separation of families

Policing—role of police, accountability for police, support for police

- Accountability, body cameras, back the Blue, defund the police, police shootings, and brutality
- Does not include prison reform, private prisons, school to prison pipeline; does not include discussions of the criminal justice system outside of the role of the police

Voting Rights—who can vote and when

- Increase the security of elections (voter ID, ballot fraud, fake results, audits, may reference January 6^{th,} but just talking about January 6th isn't enough)
- Increase access to elections (open primaries, long lines on election day, mail-in ballot, holiday)
- Also includes discussions of voting eligibility (felon enfranchisement, native enfranchisement)
- Includes discussions of changes to election rules that increase ballot access (ranked-choice voting, open primaries, nonpartisan primaries)

A3.3 Details on Machine Learning Paragraph Classifier

We train a series of supervised machine learning models to predict whether a campaign platform paragraph discusses topics related to our issue areas of interest. We train a separate classifier for each of our six issue areas (abortion, education, energy, guns, healthcare, and immigration). Following recommendations from Park and Montgomery (2023), we split our 8,306 labeled campaign platform paragraphs into an 80-20 training-validation split; this leaves us with 6,602 labeled paragraphs for model training and 1,704 labeled paragraphs for downstream validation. Before model fitting, we pre-process paragraphs by converting all text to lowercase and removing stop words and punctuation.

We train five classification models for each issue area: decision tree, gradient boosting, logistic regression, support vector machine, and random forest. For all models (other than logistic regression), we select model parameters using a 5-fold cross-validation grid search. We leverage the predictions for all five models to train a logistic regression stacking classifier. For our six classification models (five base models and one ensemble model), we make out-of-sample predictions for all unlabeled paragraphs, as well as the 1,704 paragraphs held out for downstream validation.

Tables A4 to A9 present fit statistics for our machine-learning classifiers. Across all six issue areas of interest, our ensemble classifier produces the highest out-of-sample precision, recall, and F1 Score for our validation data.

Table A4: Abortion: Out-of-Sample Validation Metrics

	Accuracy	Precision	Recall	F1
Logistic	0.975	0.954	0.338	0.500
SVM	0.989	0.940	0.758	0.839
Decision Tree	0.988	0.838	0.838	0.838
Random Forest	0.986	1.000	0.629	0.772
Gradient Boost	0.988	0.890	0.790	0.837
Stacking	0.990	0.924	0.790	0.852

Table A5: Education: Out-of-Sample Validation Metrics

	Accuracy	Precision	Recall	F1
Logistic	0.952	0.981	0.577	0.727
SVM	0.967	0.928	0.759	0.835
Decision Tree	0.957	0.827	0.770	0.797
Random Forest	0.964	0.943	0.716	0.814
Gradient Boost	0.964	0.831	0.844	0.838
Stacking	0.968	0.888	0.812	0.849

Table A6: Energy: Out-of-Sample Validation Metrics

	Accuracy	Precision	Recall	F1
Logistic	0.971	1.000	0.435	0.606
SVM	0.981	0.935	0.682	0.789
Decision Tree	0.981	0.827	0.788	0.807
Random Forest	0.977	0.979	0.564	0.716
Gradient Boost	0.981	0.827	0.788	0.807
Stacking	0.985	0.905	0.788	0.842

Table A7: Guns: Out-of-Sample Validation Metrics

	Accuracy	Precision	Recall	F1
Logistic	0.968	0.950	0.426	0.589
SVM	0.985	0.957	0.752	0.842
Decision Tree	0.985	0.911	0.808	0.857
Random Forest	0.982	0.968	0.685	0.802
Gradient Boost	0.988	0.937	0.842	0.887
Stacking	0.991	0.951	0.876	0.912

Table A8: Healthcare: Out-of-Sample Validation Metrics

	Accuracy	Precision	Recall	F1
Logistic	0.947	0.802	0.503	0.618
SVM	0.965	0.830	0.744	0.785
Decision Tree	0.955	0.721	0.786	0.752
Random Forest	0.953	0.911	0.496	0.642
Gradient Boost	0.963	0.786	0.786	0.786
Stacking	0.971	0.847	0.806	0.826

Table A9: Immigration: Out-of-Sample Validation Metrics

	Accuracy	Precision	Recall	F1
Logistic	0.955	0.948	0.433	0.594
SVM	0.980	0.960	0.771	0.855
Decision Tree	0.978	0.858	0.858	0.858
Random Forest	0.972	0.965	0.653	0.779
Gradient Boost	0.974	0.833	0.826	0.830
Stacking	0.981	0.906	0.842	0.873

A3.4 Full List of Text Pro-Processing Steps

We take the following steps when pre-processing text from our corpus of paragraphs drawn from candidates' campaign platforms:

- All text to lowercase
- String pattern 'medicare for all' to 'medicareforall'
- String pattern 'medicare-for-all' to 'medicareforall'
- String pattern 'pro choice' to 'prochoice'
- String pattern 'pro-choice' to 'prochoice'
- String pattern 'pro-life' to 'prolife'
- String pattern 'de-escalation' to 'deescalation'
- String pattern 'dodd-frank' to 'doddfrank'
- String pattern 'k-12' to 'ktwelve'
- String pattern 'k12' to 'ktwelve'
- String pattern 'pre-k' to 'prek'
- String pattern '4-year' to 'four year'
- String pattern '4 year' to 'four year'
- String pattern '2-year' to 'two year'
- String pattern '2 year' to 'two year'
- String pattern 'non-violent' to 'nonviolent'
- String pattern '2nd amendment' to 'second amendment'
- String pattern '2a' to 'second amendment'
- String pattern 'non-profit' to 'nonprofit'
- String pattern 'non-discrimination' to 'nondiscrimination'
- Replace all '-' with ' '
- Remove all non-alphabetic characters
- Remove extraneous UTC code
- Trim white space

A3.5 Doc2Vec versus Word2Vec Averaging

In the analyses below, we test the performance of CPIs generated using a different embedding model. Recall that the measurement framework presented in our main paper relies on a Doc2Vec implementation. In this implementation, embeddings are estimated for both words and collections of documents (i.e., embeddings by candidate-issue-year) simultaneously. This process is intuitively similar to taking the average of the word embeddings for candidate i on policy area j in year t. However, the Doc2Vec estimation places less weight on high-frequency words (e.g., the, is, in) with little rhetorical meaning. Previous work finds Doc2Vec outperforms averaging word embeddings (Lau and Baldwin 2016; Grimmer et al. 2022)

Employing Doc2Vec produces a superior quality of estimates. To demonstrate this, we compare CPI performance to a positioning measure produced by averaging word embeddings. Our approach for producing this alternative measure follows the exact same estimation procedure used to generate CPIs, except we estimate no candidate-issue embeddings. Instead, to produce issue-specific CPIs, we average the embeddings for all words from paragraphs identified by our classifier as covering policy area j in year t from candidate i. Hereafter, we refer to this alternative measure as "CPI-Word2Vec." We compare the performance of CPI and CPI-Word2Vec in predicting human judgments of text. We discuss our human judgment task in greater detail in Section XX, but in brief: three readers were asked to rate the positioning of platform text on a five-point scale ranging from very right (2) to very left (-2). We regress CPI and CPI-Word2Vec separately on scores produced with these human judgments and assess which of these two models (if any) better fits these human data. Results are reported in Tables A10 to A12 below. The top section of each table reports model coefficients; the bottom section reports which model best fits the data per Akaike Information Criterion (1974), a Vuong test (1989), and a Clarke test (2007). A checkmark (✓) indicates a model is preferred; for Vulong and Clarke test model differences are statistically significant at 0.05. A dash (—) indicates no model is preferred (i.e., models are indistinguishable).

Table A10: Abortion: CPI vs. CPI-Word2Vec

	Dependent variable: Human Judgement			
	Democrats		*Republicans*	
CPI	1.892*		2.599*	
	(0.512)		(0.431)	
CPI-Word2Vec		0.523		2.904*
		(0.794)		(0.585)
CFscore	−0.054	−0.079	0.201	0.283*
	(0.137)	(0.144)	(0.120)	(0.124)
Constant	−0.931*	−1.168*	0.480*	0.559*
	(0.157)	(0.170)	(0.182)	(0.189)
Observations	133	133	129	129
AIC	✓		✓	
Vuong Test	✓*		—	—
Clarke Test	✓*		✓*	
Note:				*p<0.05

Table A11: Guns: CPI vs. CPI-Word2Vec

	Dependent variable: Human Judgement			
	Democrats		*Republicans*	
CPI	4.261*		0.989*	
	(0.610)		(0.371)	
CPI-Word2Vec		6.418*		−0.026
		(1.045)		(0.517)
CFscore	−0.054	0.194	0.021	−0.015
	(0.137)	(0.211)	(0.129)	(0.133)
Constant	−0.359	−0.502*	0.940*	1.221*
	(0.216)	(0.219)	(0.197)	(0.204)
Observations	115	115	134	134
AIC	✓		✓	
Vuong Test	✓*		—	—
Clarke Test	✓*		—	—
Note:				*p<0.05

Table A12: Immigration: CPI vs. CPI-Word2Vec

	Dependent variable: Human Judgement			
	Democrats		*Republicans*	
CPI	3.542*		3.531*	
	(0.575)		(0.460)	
CPI-Word2Vec		6.816*		4.951*
		(0.124)		(0.796)
CFscore	0.232	0.251*	0.247*	0.300*
	(0.127)	(0.211)	(0.094)	(0.100)
Constant	−0.302*	0.001	0.275*	0.488*
	(0.216)	(0.158)	(0.129)	(0.131)
Observations	128	128	127	127
AIC		✓	✓	
Vuong Test	—	—	✓*	
Clarke Test	✓*		✓*	
Note:				*p<0.05

A3.6 Term Selection Strategy & Robustness

To determine the terms that will define the left-most and right-most positions for our eight issue cleavages, we first consulted the policy platforms of far-left and far-right advocacy groups. We additionally consult policy priorities made available by far-left and far-right congressional caucuses. For each issue area, we consulted the following resources:

Abortion: Far-Right

- Eagle Forum: https://eagleforum.org/topics/pro-life.html
- Family Research Council: https://www.frc.org/abortion
- Heritage Foundation: https://www.heritage.org/life-and-family
- National Right to Life: https://nrlc.org/
- Susan B. Anthony: https://sbaprolife.org/about

Abortion: Far-Left

- Center for American Progress: https://www.americanprogress.org/topic/abortion-rights/
- House Progressive Caucus: https://progressives.house.gov/universal-health-care
- Justice Democrats: https://justicedemocrats.com/platform/society/#reproductive-rights
- Our Revolution: https://ourrevolution.com/policy-fights/
- Planned Parenthood: https://www.plannedparenthoodaction.org/issues/abortion
- Reproductive Freedom for All: https://reproductivefreedomforall.org/about/

Education: Far-Right

- Club for Growth: https://www.clubforgrowth.org/issue/education/
- Conservative Caucus: https://www.theconservativecaucus.org/about/stand-for-fight-for
- CPAC: http://ratings.conservative.org/issues?group=B
- Eagle Forum: https://eagleforum.org/topics/education.html
- Family Research Council: https://www.frc.org/education
- FreedomWorks: https://www.freedomworks.org/issue/curriculum/
- Heritage Foundation: https://www.heritage.org/empower-parents-make-education-choices

Education: Far-Left

- Center for American Progress: https://www.americanprogress.org/topic/education-k-12/
- Congressional Progressive Caucus Center: https://www.progressivecaucuscenter.org/debt-free-college
- House Progressive Caucus: https://progressives.house.gov/education

- Justice Democrats: https://justicedemocrats.com/platform/economy/#cancel-student-debt

Energy: Far-Right

- CPAC: http://ratings.conservative.org/issues?group=C
- Conservative Caucus: https://www.theconservativecaucus.org/about/stand-for-fight-for
- FreedomWorks: https://www.freedomworks.org/vote/energy-independence-and-security-act/
- Heritage Foundation: https://www.heritage.org/energy

Energy: Far-Left

- Center for American Progress: https://www.americanprogress.org/topic/clean-energy-2/
- Congressional Progressive Caucus Center: https://www.progressivecaucuscenter.org/issues/environment-climate
- House Progressive Caucus: https://progressives.house.gov/climate-justice
- Our Revolution: https://ourrevolution.com/policy-fights/

Guns: Far-Right

- CPAC: http://ratings.conservative.org/issues?group=A
- Gun Owners of America: https://www.gunowners.org/about-goa/
- National Association for Gun Rights: https://www.nationalgunrights.org/about-us/nagr-pac/
- National Rifle Association: https://home.nra.org/statements/nra-statement-on-gun-control-package/

Guns: Far-Left

- Center for American Progress: https://www.americanprogress.org/topic/gvp/
- Congressional Progressive Caucus Center: https://www.progressivecaucuscenter.org/issues/gun-violence-prevention
- Everytown for Gun Safety: https://www.everytown.org/
- Justice Democrats: https://justicedemocrats.com/platform/society/#gun-safety
- Moms Against Gun Violence: https://momsdemandaction.org/about/

Healthcare: Far-Right

- CPAC: http://ratings.conservative.org/issues?group=F
- Family Research Council: https://www.frc.org/health-care

Healthcare: Far-Left

- Center for American Progress: https://www.americanprogress.org/topic/health-coverage-and-access/
- House Progressive Caucus: https://progressives.house.gov/universal-health-care
- Medicare-For-All Action Network https://actionnetwork.org/letters/medicare-for-all-caucus/
- Our Revolution: https://ourrevolution.com/policy-fights/
- Congressional Progressive Caucus Center: https://www.progressivecaucuscenter.org/medicare-for-all

Immigration: Far-Right

- Conservative Caucus: https://www.theconservativecaucus.org/about/stand-for-fight-for
- Heritage Foundation: https://www.heritage.org/borders-and-crime

Immigration: Far-Left

- Center for American Progress: https://www.americanprogress.org/topic/immigration/
- Congressional Progressive Caucus Center: https://www.progressivecaucuscenter.org/issues/immigrants-rights
- House Progressive Caucus: https://progressives.house.gov/immigrant-rights

We principally relied on these resources to produce our dictionaries of terms, provided in Table A11 below. In each dictionary, we include at least 25 terms to ensure that our measurement results are not term-dependent. Extant work shows that long term lists well-capture the semantic meaning of a latent construct of interest. As a final step, we check whether our corpus of text reflects the concepts we seek to capture in our positional dictionaries. We do this by cross-referencing our dictionaries of terms with the vocabulary of our corpus; we then randomly sample documents to make sure that terms are used in an expected context. An alternative approach to term selection could rely on a computer-assisted, statistical approach to keyword selection. When implementing the approach proposed by King et al. (2017) for computer-assisted keyword discovery, we found our term list more closely reflected our core concepts of interest.

Table A13: Issue Cleavages & Full Positional Term Dictionaries

Issue Area	Left-Most Position	Right-Most Position
Abortion	**Pro-Choice**: reproductive, justice, freedom, health, care, healthcare, expand, access, safe, hyde, roe, wade, choice, prochoice, government, codifying, legal, restrictive, decisions, autonomy, control, privacy, doctor, respect, personal, equality, women, sexual, services	**Pro-Life:** life, birth, death, unborn, womb, precious, sanctity, moment, conception, heartbeat, begins, prolife, partial, fetal, born, alive, demand, saving, defund, taxpayer, infanticide, murder, barbaric, innocent, dignity, values, conscientious, immoral, vulnerable, defenseless, ban, outlaw, god, gift, sacred
Education	**Federal Involvement:** public, underfunded, invest, government, responsibility, cuts, investment, expand, teachers, affordable, free, tuition, debt, universal, profit, prek, loan, cancel, college, forgiveness, income, race, live, zip, accessible, equality, pay, salaries, technology, head, start	**Local Control:** parents, bureaucrats, burdensome, tax, state, local, board, know, best, values, control, autonomy, decisions, abolish, eliminate, dismantle, department, decentralize, choice, competition, religious, charter, private, homeschool, vouchers, mandates, core, common, curriculum
Energy	**Renewables Investment:** renewable, clean, green, sustainable, solar, wind, hydroelectric, transition, paris, rejoin, emissions, climate, change, environment, ban, pollution, technology, research, development, incentives, credits, tax, subsidize, funding, investment, new, deal, net, zero, hundred	**Fossil Fuel Investment:** oil, gas, coal, keystone, pipeline, independence, foreign, reliance, deregulate, cut, reopen, restrictions, red, tape, bureaucratic, repeal, free, market, private, competition, companies, china, russia, middle, east, national, security, disruption, unproven, unstable, domestic, expand, production, reserves
Guns	**Increase Restrictions:** mandatory, background, ban, national, hate, registry, database, assault, automatic, ar, ak, military, battlefield, war, waiting, age, years, ghost, loopholes, transfer, traumatic, violence, mass, killing, crisis, epidemic, dealers, manufacturers, trafficking, comprehensive, stricter	**Increase Access:** second, amendment, owner, abiding, right, infringed, inherent, unconstitutional, founding, enumerated, confiscation, unrestricted, repeal, oppose, abolish, any, overturn, enshrined, reciprocity, concealed, carry, hearing, security, freedoms, fundamental, defending, families, protection

Health Insurance	**Publicly Funded:** universal, all, deny, coverage, privilege, guarantee, uninsured, gap, public, medicareforall, singlepayer, free, strengthen, expand, comprehensive, dental, hearing, vision, access, human, right, equality, fundamental, poor, justice, inclusive, profit, preexisting, bankruptcy	**Privatize:** repeal, replace, problems, flawed, free, market, open, competition, freedom, national, state, choice, consumers, individual, patient, provider, insurance, mandate, Washington, burdensome, tape, involvement, bureaucratic, deregulate, tax, hsa, savings, obamacare, socialized
Immigration	**Inclusive:** undocumented, pathway, roadmap, temporary, tps, dreamers, daca, asylum, humane, compassion, dignity, diverse, culture, xenophobic, racist, families, reunification, detention, cages, humanitarian, streamline, backlogs, courts, accessible, opportunity, expand, abolish, ice, education, healthcare	**Exclusive:** enforce, rule, law, secure, build, wall, barrier, surveillance, technology, verify, screening, safety, protect, sanctuary, chain, migration, anchor, birthright, english, skilled, merit, labor, overstay, deport, amnesty, illegal, criminal, terrorists, flooding, porous, close, funding, hire, officers, patrol

To demonstrate that our estimates are robust to alternative term selections, we randomly sample terms from Table A11 and re-estimate CPIs. We repeat this process a total of 50 times for each issue area. We then correlate each of these scores with the CPIs used in the body of the paper. Consistently, we find very high correlations with our measures, both across and within party. Below are the average correlations across the 50 keyword samples for each issue area. This analysis demonstrates that CPI estimation is not sensitive to excluding or including specific keywords.

- Abortion: Cross-Party (0.983); Dems (0.920); Reps (0.939)
- Education: Cross-Party (0.938); Dems (0.866); Reps (0.919)
- Energy: Cross-Party (0.944); Dems (0.899); Reps (0.933)
- Guns: Cross-Party (0.985); Dems (0.941); Reps (0.967)
- Healthcare: Cross-Party (0.953); Dems (0.900); Reps (0.923)
- Immigration: Cross-Party (0.942); Dems (0.882); Reps (0.885)

A4 Measurement Validation

A4.1 Coding Instructions for Human Judgment Task

Coding Abortion Policy Statements (Pro-Choice vs. Pro-Life)

This task involves reading text from congressional candidates' campaign platforms and judging how much they are left-leaning (very left, left) or right-leaning (very right, right). If you believe the text expresses a centrist position, select the "centrist" option. If you believe the text does not express a clear position, select the "ambiguous" option. If you believe the text does not express a position relevant to abortion policy, select the "irrelevant" option.

What are "left" or "right" abortion policies?

Very left positions tend to advocate for the following:

- Access to abortion without restrictions, removing any and all obstacles
- Safeguarding abortion rights through **federal legislation**
  - Codifying Roe vs. Wade
  - Repeal the Hyde Amendment, global gag rule
- Views abortion as a **fundamental right of women**
  - "Abortions are healthcare", "reproductive justice", "reproductive rights"

Left positions tend to advocate for the following:

- **General pro-choice rhetoric** without explicit discussions of putting protections into law
- Views abortion as a choice that should be available to women
  - "Pro-choice", "Protecting women's right to choose"
- May express that they personally do not believe in abortion, but still should be available

Right positions tend to advocate for the following:

- **General pro-life rhetoric**, abortions being unlawful in nearly all cases
  - Exceptions include **rape, incest, child pregnancy**
- Simply states they are "pro-life", "right to life", or "life is precious"

Very right positions tend to advocate for the following:

- Explicitly says **no abortions** under any circumstances **(only exception is the life of the mother)**
- Views abortion as murder/homicide
- Outlawing abortion through federal legislation
- Includes rhetoric like: **"Life begins at conception"**

What is a centrist abortion policy?

- Abortions **available to the general public with restrictions** (limit to first trimester, **20 weeks**)
- Takes an equivalent proportion of left and right positions

What is an ambiguous abortion policy?

- Discusses abortion in terms of reproductive education, and adoption options without taking an explicit pro-life/pro-choice position on the issue
- Discusses reproductive healthcare without reference to bodily choices

What is an irrelevant abortion policy?

- Discusses reproduction (e.g., access to contraception, education), women's healthcare, or Planned Parenthood without any discussion of abortion

Coding Education Policy Statements (Federal vs. Local)

This task involves reading text from congressional candidates' campaign platforms and judging how much they are left-leaning (very left, left) or right-leaning (very right, right). If you believe the text expresses a centrist position, select the "centrist" option. If you believe the paragraph does not express a clear position, select the "ambiguous" option. If you believe the text does not express a position relevant to immigration policy, select the "irrelevant" option.

What are "left" or "right" education policies?

Very left positions tend to advocate for the following:

- Education is made better *with **maximal government intervention***
- Left positions *plus some combination of:*
    - **Universal pre-k, free college for all, college for all**
    - **Public college free, community college free**

Left positions tend to advocate for the following:

- Education is made better *with **more government intervention***
    - Prioritizing **funding for public schools**, tax incentives for vocational programs, making **college more affordable**, providing some loan forgiveness/**better lending programs,** giving **public school teachers the resources**/pay they need, increasing federal funding for schools, creating more **equity across public schools, reducing the cost of college**

Right positions tend to advocate for the following:

- Education is made better *with **less government intervention***
    - Curriculum choice, **parent choice over schools**, homeschooling, more options beyond public school, remove Common Core, No Child Left Behind

Very right positions tend to advocate for the following:

- Education is made better *with **minimal government intervention***
- Right positions *plus some combination of:*
    - **Disband/weaken the Department of Education**; remove the federal government entirely
    - Put **all control in the hands of parents**, local school boards, and state government
    - **End all federal funding**

What is a centrist education policy?

- A broad mix of both left and right; **public/private partnership**

What is an ambiguous education policy?

- We need to "improve" education without a clear policy proposal
- Discussion of improving education, quality, but no discussion of government role

What is an irrelevant education policy?

- Discussions of school-to-prison pipeline, STEM, curriculum (beyond Common Core, NCLB)

Coding Energy Policy Statements (Renewables vs. Fossil Fuels)

This task involves reading text from congressional candidates' campaign platforms and judging how much they are left-leaning (very left, left) or right-leaning (very right, right). If you believe the text expresses a centrist position, select the "centrist" option. If you believe the paragraph does not express a clear position, select the "ambiguous" option. If you believe the text does not express a position relevant to immigration policy, select the "irrelevant" option.

What are "left" or "right" energy policies?

Very left positions tend to advocate for the following:

- **Full investment in renewable energy** (I am 100% committed…)
- Pass legislation to provide tax incentives, **ban fossils, subsidize green energy growth**
    - Rejoin Paris Climate Agreement; Green New Deal; net zero emissions by…

Left positions tend to advocate for the following:

- More in favor of renewable energy
- In favor of **increasing renewable energy,** committed to growing green energy
- Broad statements "I support" statements

Right positions tend to advocate for the following:

- More **in favor of fossil energy; market-driven energy policy**
- Protect the fossil energy sector; "we need fossil energy"
    - Reopen pipelines/fossil fuel facilities, open keystone pipeline
    - May discuss green energy as an unsuitable alternative (capacity, unproven, expensive)

Very right positions tend to advocate for the following (fossil, free market most important)

- **Fully investment in fossil fuels** (I am 100% committed…)
- **Remove regulations, abolish EPA, reduce government regulations on fossil fuels**

What is a centrist energy policy?

- "**All of the above approach**" to energy policy
- Discusses both fossil and renewable energy equally
- Says we must "balance" environmental and economic considerations
- Note: natural gas *is not* renewable

What is an ambiguous energy policy?

- Discusses mechanisms of technology (e.g., nuclear energy is safe, natural gas is clean…)
- We need more energy production; we need to invest in energy production

What is an irrelevant energy policy?

- Discussions of climate change *only*; environmental conservation *only*

Coding Gun Policy Statements (Restrition vs. Access)

This task involves reading text from congressional candidates' campaign platforms and judging how much they are left-leaning (very left, somewhat left) or right-leaning (very right, somewhat right). If you believe the text expresses a centrist position, select the "centrist" option. If you believe the paragraph does not express a clear position, select the "ambiguous" option. If you believe the text does not express a position relevant to immigration policy, select the "irrelevant" option.

What are "left" or "right" gun policies?

Very left positions tend to advocate for the following:

- **Purchase restrictions on guns** themselves (military-style, assault rifles, AK-47s, weapons of war)
- Repeal legal protections for **gun manufacturers, dealers**
- Ban all weapons

Left positions tend to advocate for the following:

- Some restrictions on **buying time; wait periods; raising the age of purchase**
    - When and where you can purchase (gun shows, wait periods, private sales)
- **Special permitting** for assault-style, certain weapons
- Creation of a national **registry** of gun owners
- Generally, make it harder to purchase guns, "stricter gun laws"

Right positions tend to advocate for the following:

- Discussion of support for 2nd amendment, but **no discussion of restrictions**
- Prevent liberals from taking guns away
- Against left gun positions (restrictions on where, when, and what can be bought)

Very right positions tend to advocate for the following:

- Repealing limitations on concealed carry, other gun access laws
    - **Reciprocity** in concealed carry across state lines
- Guns in schools, and college campuses (i.e., **"gun-free" zones**), provide teachers with guns
- No background checks, no waiting periods, no classes/permitting
- **Against centrist positions (basic, common-sense reforms)**
- No limits on guns whatsoever; **ANY gun control laws are a problem**

What is a centrist gun policy?

- **No-fly lists; red flag laws; mental health; closing loopholes; universal background checks**
- "Common sense" gun reforms
- **Ban on bump stocks, high-capacity magazines**
- Buy-back programs, with the option of keeping guns
- Can be pro 2nd Amendment, but mentions restrictions

What is an ambiguous gun policy?

- Discussions of responsible gun ownership; common-sense measures
- "Get guns off our streets", and "reduce gun violence" with no policy discussion

What is an irrelevant gun policy?

- Discussions of the prison system, cash bail

Coding Health Insurance Marketplace Statements (Public vs. Private)

This task involves reading text from congressional candidates' campaign platforms and judging how much they are left-leaning (very left, left) or right-leaning (very right, right). If you believe the text expresses a centrist position, select the "centrist" option. If you believe the text does not express a clear position, select the "ambiguous" option. If you believe the text does not express a position relevant to healthcare policy, select the "irrelevant" option.

What are "left" or "right" healthcare policies?

Very left positions tend to advocate for the following:

- **Full government involvement** in the health insurance market
    - **Single-payer, universal healthcare system, "Medicare-for-all"; abolish private insurance**
- Healthcare is a fundamental human right

Left positions tend to advocate for the following:

- **Expanded government involvement** in the health insurance market:
    - **Expand public options**; **expand Medicaid; expand Medicare**
    - May advocate for keeping private insurance for those who want it

Right positions tend to advocate for the following:

- **Little government involvement** in the health insurance market:
    - **Ensure preexisting condition coverage, cost caps on insurance**
- Advocates for choice on coverage but **does not go as far as to say no government involvement**
- **Free market**, purchasing across state lines, competition

Very right positions tend to advocate for the following:

- Remove government from health insurance marketplace, full free market
- **Repeal Obamacare; ACA; replace with competitive, free-market system**

What is a centrist healthcare policy?

- Maintain status quo; **protect** ACA; **protect** Medicare; **protect** Medicaid

What is an ambiguous healthcare policy?

- Broadly discusses reform to achieve affordability, lower cost, and greater access but **takes no explicit position**

What is an irrelevant healthcare policy?

- Discusses ONLY quality of healthcare, scope of services for individual populations, group-based
    - Black maternal mortality, veteran's healthcare, women's healthcare
    - Drug costs

Coding Immigration Policy Statements (Inclusive vs. Exclusive)

This task involves reading text from congressional candidates' campaign platforms and judging how much they are left-leaning (very left, somewhat left) or right-leaning (very right, somewhat right). If you believe the text expresses a centrist position, select the "centrist" option. If you believe the paragraph does not express a clear position, select the "ambiguous" option. If you believe the text does not express a position relevant to immigration policy, select the "irrelevant" option.

What are "left" or "right" immigration policies?

Very left positions tend to advocate for the following:

- Treatment of immigrants: **pathway to citizenship for *all* undocumented**
- Calls to end all deportation
    - Provide healthcare, education, humane treatment
- Border Control, Security: **Abolish, demilitarize ICE**
- General Approach: inclusive; maximize access to immigration

Left positions tend to advocate for the following:

- Treatment of immigrants**: some pathway** to citizenship (e.g., **DACA only**)
    - Earned pathway: military service, civil service; provide temporary protected status (TPS)
- Border Control, Security: **No kids in cages, humane treatment** at border
- General Approach: inclusive; expand access to immigration
    - A mix of law + citizenship, with **more emphasis on pathways to citizenship**

Right positions tend to advocate for the following:

- Treatment of immigrants: **emphasize lawful immigration**; illegal immigrants must go through the regular immigration process, with no special treatment
- Recognize the need for seasonal, temporary workers; keeping "talented" illegal immigrants
- Border Control: **Secure the border;** investment in security
- General Approach: exclusive; reduce access to immigration
    - A mix of law + citizenship, with **more emphasis on security**

Very right positions tend to advocate for the following:

- Treatment of immigrants: **all illegal immigrants as criminals; no amnesty; deport all illegal**
- increased limits on who can immigrate (support for Muslim ban, only high-skilled workers, country-specific restrictions); **end birthright citizenship; end chain migration**
- Border Control: Heavy investment in border security
- **Close borders, end all immigration**

What is a centrist immigration policy?

- An **even mix of enforcing laws at the border + some pathway to citizenship**

What is an ambiguous immigration policy?

- Just platitudes: we are a nation of immigrants, treat everyone with respect

What is an irrelevant immigration policy?

- None

A4.2 Coding Descriptive Statistics

Below, we explore descriptive statistics from our human judgment task. Table A14 outlines the coder agreement in categorizing policy platform paragraphs. For most texts, at least two readers agreed on a common score for a given policy platform. Table A15 provides the proportion of documents in each issue area flagged for including only irrelevant text; guidance on coding for irrelevant text varied by issue area and is outlined in greater detail in Section A4.1. On average, readers flagged about 3% or ten documents per issue area is including only irrelevant text. Table A16 provides the proportion of documents in each issue area flagged for taking an ambiguous position; guidance on coding for ambiguous positioning text varied by issue area and is outlined in greater detail in Section A4.1. On average, readers flagged about 5% or fifteen documents per issue area as including text that was ambiguous in position-taking content .

Table A14: Reader Agreement Rate, By Issue Area

Issue Area	3/3 Agreement	2/3 Agreement	No Agreement
Abortion	0.64	0.31	0.04
Guns	0.54	0.41	0.04
Immigration	0.45	0.51	0.04

Table A15: Flagged Irrelevant Text, by Issue Area

Issue Area	Reader A	Reader B	Reader C	Reader D
Abortion	0.01	0.07	0.03	—
Guns	0.01	—	0.01	0.01
Immigration	0.03	0.05	—	0.03

Table A16: Flagged Ambiguous Text, by Issue Area

Issue Area	Reader A	Reader B	Reader C	Reader D
Abortion	0.02	0.02	0.03	—
Guns	0.03	—	0.06	0.05
Immigration	0.09	0.10	—	0.05

A4.3 Full Validation Model Outputs & Alternative Specifications

Table A17: CPI Validation with Human Judgements: Ordinal Logit

	Dependent variable: Human Judgement					
	All		*Democrats*		*Republicans*	
CPI-Abortion	9.420*	7.801*	8.067*	5.937*	10.583*	9.360*
	(1.392)	(1.487)	(2.020)	(2.099)	(1.914)	(2.148)
DIME-CFScore		0.021		−0.802		0.657
		(0.378)		(0.563)		(0.509
Candidate Party	4.199*	5.786*				
	(0.744)	(1.434)				
Observations	289	262	137	133	156	129
CPI-Guns	6.610*	6.570*	11.614*	11.934*	3.953*	3.138
	(1.091)	(1.273)	(2.076)	(2.288)	(1.350)	(1.645)
DIME-CFScore		0.598		1.007		0.278
		(0.416)		(0.637)		(0.552)
Candidate Party	4.727*	4.153*				
	(0.789)	(1.414)				
Observations	289	249	122	115	167	134
CPI-Immigration	12.832*	12.000*	10.910*	10.175*	12.911*	12.586*
	(1.536)	(1.663)	(2.093)	(2.218)	(2.288)	(2.552)
DIME-CFScore		1.230*		1.382*		1.372*
		(0.322)		(0.602)		(0.438)
Candidate Party	3.183*	1.707				
	(0.578)	(0.931)				
Observations	296	255	142	128	154	127

Table A18: CPI Validation with Human Judgements: Web Ideology

	Dependent variable: Human Judgement					
	All		*Democrats*		*Republicans*	
CPI-Abortion	2.964*	2.282*	2.816*	1.964*	3.077*	2.399*
	(0.308)	(0.413)	(0.489)	(0.677)	(0.393)	(0.541)
Case-WEB		0.366*		0.524*		0.315*
		(0.107)		(0.223)		(0.120)
Candidate Party	1.381*	0.963*				
	(0.131)	(0.193)				
Constant	−0.716*	−0.445*	−0.737*	−0.362*	0.638*	0.537*
	(0.062)	(0.096)	(0.082)	(0.156)	(0.101)	(0.140)
Observations	289	289	137	137	156	156
CPI-Guns	2.513*	2.487*	4.377*	3.851*	1.292*	0.870
	(0.309)	(0.453)	(0.547)	(0.689)	(0.331)	(0.536)
Case-WEB		0.122		0.367		0.111
		(0.139)		(0.268)		(0.140)
Candidate Party	1.350*	1.248*				
	(0.131)	(0.224)				
Constant	−0.732*	−0.711*	−0.495*	−0.366	0.888*	0.899*
	(0.069)	(0.114)	(0.094)	(0.193)	(0.084)	(0.131)
Observations	289	289	122	122	167	167
CPI-Immigration	4.048*	3.355*	3.896*	3.448*	4.176*	3.256*
	(0.354)	(0.455)	(0.561)	(0.720)	(0.449)	(0.583)
Case-WEB		0.366*		0.396*		0.341*
		(0.103)		(0.166)		(0.130)
Candidate Party	0.943*	0.621*				
	(0.102)	(0.156)				
Constant	−0.410*	−0.248*	−0.429*	−0.214	0.521*	0.403*
	(0.064)	(0.085)	(0.086)	(0.116)	(0.060)	(0.105)
Observations	296	296	142	142	154	154

Table A19: CPI Validation with Human Judgements: Main Paper Analysis

	Dependent variable: Human Judgement					
	All		*Democrats*		*Republicans*	
CPI-Abortion	2.964*	2.315*	2.816*	1.892*	3.077*	2.599*
	(0.308)	(0.332)	(0.489)	(0.512)	(0.393)	(0.431)
DIME-CFScore		0.089		−0.054		0.201
		(0.090)		(0.137)		(0.120)
Candidate Party	1.381*	1.431*				
	(0.131)	(0.243)				
Constant	−0.716*	−0.738*	−0.737*	−0.931*	0.638*	0.480*
	(0.062)	(0.105)	(0.082)	(0.157)	(0.101)	(0.182)
Observations	289	262	137	133	156	129
CPI-Guns	2.513*	2.385*	4.377*	4.261*	1.292*	0.989*
	(0.309)	(0.354)	(0.547)	(0.610)	(0.331)	(0.371)
DIME-CFScore		0.116		0.190		0.021
		(0.121)		(0.204)		(0.129)
Candidate Party	1.350*	1.174*				
	(0.131)	(0.313)				
Constant	−0.732*	−0.680*	−0.495*	−0.359	0.888*	0.940*
	(0.069)	(0.137)	(0.094)	(0.216)	(0.084)	(0.197)
Observations	289	249	122	115	167	134
CPI-Immigration	4.048*	3.536*	3.896*	3.542*	4.176*	3.531*
	(0.354)	(0.364)	(0.561)	(0.575)	(0.449)	(0.460)
DIME-CFScore		0.241*		0.232		0.247*
		(0.077)		(0.127)		(0.094)
Candidate Party	0.943*	0.578*				
	(0.102)	(0.190)				
Constant	−0.410*	−0.295*	−0.429*	−0.302*	0.521*	0.275*
	(0.064)	(0.093)	(0.086)	(0.139)	(0.060)	(0.129)
Observations	296	255	142	128	154	127

Note: *p<0.05

Table A20: CPI Validation with Pro-Choice PAC Giving: Democratic Candidates

	(Main 1)	(Main 2)	(Alternative 1)	(Alternative 2)
CPI-Abortion	-0.368^{*}	-0.284^{*}		
	(0.077)	(0.081)		
CPI-Abortion (Imputed at party mean)			-0.326^{*}	
			(0.083)	
CPI-Abortion (Imputed at moderate)				-0.199^{*}
				(0.039)
DIME-CFscore		1.373^{*}	0.845^{*}	0.922^{*}
		(0.263)	(0.121)	(0.123)
Constant	-1.404^{*}	0.113	-0.643^{*}	-0.100
	(0.171)	(0.308)	(0.201)	(0.116)
Observations	1,030	973	2,397	2,397

Note: CFscores range from negative (liberal) to positive (conservative) values. CFscore coefficients here are positive, indicating that as a candidate becomes more conservative, they are *more likely* to receive pro-choice PAC funding. Running a bivariate model with CFscore and PAC giving produces the same counter-intuitive result. $p < 0.05$.

Table A21: CPI Validation with Pro-Life PAC Giving: Republican Candidates

	(Main 1)	(Main 2)	(Alternative 1)	(Alternative 2)
CPI-Abortion	0.191^{*}	0.140^{*}		
	(0.065)	(0.071)		
CPI-Abortion (imputed at party mean)			0.139	
			(0.072)	
CPI-Abortion (imputed at moderate)				0.185^{*}
				(0.041)
DIME-CFscore		-0.848^{*}	-0.332^{*}	-0.386^{*}
		(0.193)	(0.098)	(0.100)
Constant	-1.071^{*}	0.347	-0.523^{*}	-0.359^{*}
	(0.134)	(0.272)	(0.174)	(0.128)
Observations	1,134	955	2,226	2,226

Note: CFscores range from negative (liberal) to positive (conservative) values. CFscore coefficients here are negative, indicating that as a candidate becomes more conservative, they are *less likely* to receive pro-choice PAC funding. Running a bivariate model with CFscore and PAC giving produces the same counter-intuitive result. $p < 0.05$.

Table A22: CPI Validation with Pro-Gun Control PAC Giving: Democratic Candidates

	(Main 1)	(Main 2)	(Alternative 1)	(Alternative 2)
CPI-Guns	-0.220^{*}	-0.180^{*}		
	(0.073)	(0.075)		
CPI-Guns (imputed at party mean)			-0.199^{*}	
			(0.077)	
CPI-Guns (imputed at moderate)				-0.138^{*}
				(0.043)
DIME-CFscore		0.391	0.607^{*}	0.632^{*}
		(0.223)	(0.130)	(0.130)
Constant	-1.875^{*}	-1.352^{*}	-1.305^{*}	-1.033^{*}
	(0.163)	(0.257)	(0.187)	(0.125)
Observations	1,162	1,072	2,397	2,397

Note: CFscores range from negative (liberal) to positive (conservative) values. CFscore coefficients here are positive, indicating that as a candidate becomes more conservative, they are *more likely* to receive pro-gun control PAC funding. Running a bivariate model with CFscore and PAC giving produces the same counter-intuitive result. $p < 0.05$.

Table A23: CPI Validation with Pro-Gun Rights PAC Giving: Republican Candidates

	(Main 1)	(Main 2)	(Alternative 1)	(Alternative 2)
CPI-Guns	0.329^{*}	0.265^{*}		
	(0.051)	(0.056)		
CPI-Guns (imputed at party mean)			0.278^{*}	
			(0.056)	
CPI-Guns (imputed at moderate)				0.218^{*}
				(0.038)
DIME-CFscore		-0.685^{*}	-0.440^{*}	-0.470^{*}
		(0.154)	(0.100)	(0.100)
Constant	-1.390^{*}	-0.160	-0.610^{*}	-0.318^{**}
	(0.108)	(0.213)	(0.151)	(0.129)
Observations	1,396	1,143	2,226	2,226

Note: CFscores range from negative (liberal) to positive (conservative) values. CFscore coefficients here are negative, indicating that as a candidate becomes more conservative, they are *less likely* to receive pro-gun rights PAC funding. Running a bivariate model with CFscore and PAC giving produces the same counter-intuitive result. $p < 0.05$.

A5 Measurement Application

Figure A9: Candidate Positioning Index Party Mean, By Issue and Year

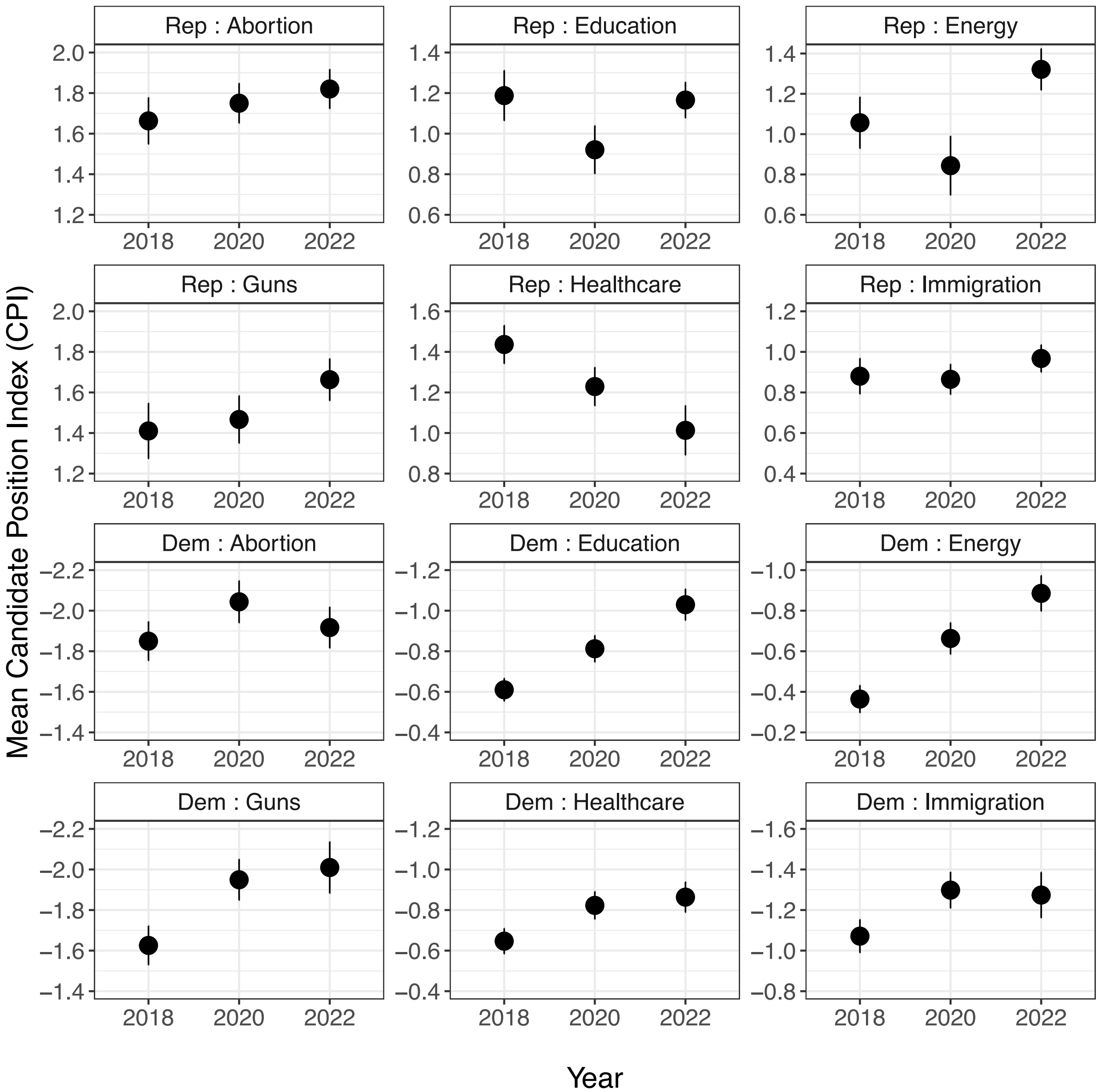


Table A24: Candidate Responsiveness to District Opinion: Abortion

	Dependent variable: CPI-Abortion			
	(CPI)	(CPI)	(CFScore)	(CFScore)
District Op.: Left Policy	-0.736^{*}		-0.221	
	(0.269)		(0.171)	
District Op.: Right Policy		1.007^{*}		0.255
		(0.371)		(0.235)
Republican Candidate	1.756^{*}	1.757^{*}	1.998^{*}	1.999^{*}
	(0.024)	(0.024)	(0.015)	(0.015)
Dem. Safe	0.002	-0.003	0.125^{*}	0.121^{*}
	(0.035)	(0.034)	(0.022)	(0.021)
Rep. Safe	0.053	0.063^{*}	-0.090^{*}	-0.085^{*}
	(0.034)	(0.032)	(0.021)	(0.021)
Prev. Elected Exp.	-0.020	-0.021	0.011	0.010
	(0.030)	(0.030)	(0.019)	(0.019)
Incumbent	-0.053^{*}	-0.050^{*}	0.016	0.017
	(0.027)	(0.027)	(0.017)	(0.017)
Non-Male	-0.038	-0.038	-0.029^{*}	-0.030^{*}
	(0.023)	(0.023)	(0.015)	(0.015)
Open	-0.047^{*}	-0.046^{*}	0.013	0.013
	(0.028)	(0.028)	(0.018)	(0.018)
FE: Year-State	✓	✓	✓	✓
Constant	-0.562^{*}	-1.484^{*}	-0.896^{*}	-1.149^{*}
	(0.201)	(0.229)	(0.127)	(0.145)
Observations	1,927	1,927	1,927	1,927
R^2	0.826	0.826	0.932	0.932
Note:				$^{*}p<0.05$

Table A25: Candidate Responsiveness to District Opinion: Guns

	Dependent variable: CPI-Guns			
	(CPI)	(CPI)	(CFScore)	(CFScore)
District Op.: Left Policy	−0.157		0.234	
	(0.337)		(0.202)	
District Op.: Right Policy		0.406		−0.278
		(0.357)		(0.214)
Republican Candidate	1.621*	1.619*	1.946*	1.946*
	(0.026)	(0.026)	(0.016)	(0.016)
Dem. Safe	−0.092*	−0.085*	0.092*	0.094*
	(0.039)	(0.038)	(0.024)	(0.023)
Rep. Safe	0.154*	0.143*	−0.086*	−0.085*
	(0.038)	(0.037)	(0.023)	(0.022)
Prev. Elected Exp.	0.045	0.044	0.031	0.031
	(0.036)	(0.036)	(0.021)	(0.021)
Incumbent	0.014	0.014	0.025	0.024
	(0.030)	(0.030)	(0.018)	(0.018)
Non-Male	−0.008	−0.007	−0.045*	−0.045*
	(0.026)	(0.027)	(0.016)	(0.016)
Open	−0.018	−0.020	0.026	0.027
	(0.032)	(0.032)	(0.019)	(0.019)
FE: Year-State	✓	✓	✓	✓
Constant	−0.775*	−1.016*	−1.191*	−0.945*
	(0.259)	(0.194)	(0.155)	(0.116)
Observations	2,214	2,214	2,214	2,214
R^2	0.742	0.742	0.908	0.908
Note:				*p<0.05

Table A26: Candidate Responsiveness to District Opinion: Healthcare

	Dependent variable: CPI-Healthcare			
	(CPI)	(CPI)	(CFScores)	(CFScores)
District Op.: Left Policy	-1.134^{*}		0.248	
	(0.447)		(0.188)	
District Op.: Right Policy		0.622^{*}		0.060
		(0.375)		(0.157)
Republican Candidate	1.379^{*}	1.381^{*}	1.865^{*}	1.864^{*}
	(0.026)	(0.026)	(0.011)	(0.011)
Dem. Safe	-0.031	-0.060	0.107^{*}	0.121^{*}
	(0.041)	(0.039)	(0.018)	(0.016)
Rep. Safe	0.045	0.060	-0.080^{*}	-0.091^{*}
	(0.039)	(0.039)	(0.017)	(0.017)
Prev. Elected Exp.	0.001	0.001	0.055^{*}	0.055^{*}
	(0.039)	(0.039)	(0.015)	(0.015)
Incumbent	0.216^{*}	0.215^{*}	0.065^{*}	0.064^{*}
	(0.031)	(0.031)	(0.013)	(0.013)
Non-Male	-0.171^{*}	-0.172^{*}	-0.041^{*}	-0.040^{*}
	(0.027)	(0.027)	(0.012)	(0.012)
Open	-0.014	-0.012	0.021	0.021
	(0.034)	(0.034)	(0.014)	(0.014)
FE: Year-State	✓	✓	✓	✓
Constant	0.540	-0.555^{*}	-1.161^{*}	-1.010^{*}
	(0.378)	(0.261)	(0.156)	(0.104)
Observations	2,795	2,795	2,795	2,795
R^2	0.585	0.585	0.890	0.890
Note:				*p<0.05

Table A27: Candidate Responsiveness to District Opinion: Immigration

	Dependent variable: CPI-Immigration			
	(CPI)	(CPI)	(CFScore)	(CFScore)
District Op.: Left Policy	−2.672*		0.128	
	(0.706)		(0.382)	
District Op.: Right Policy		1.303*		−0.411*
		(0.455)		(0.241)
Republican Candidate	1.479*	1.445*	1.960*	1.919*
	(0.035)	(0.029)	(0.019)	(0.015)
Dem. Safe	−0.120*	−0.142*	0.067*	0.085*
	(0.053)	(0.041)	(0.029)	(0.022)
Rep. Safe	−0.020	0.045	−0.074*	−0.073*
	(0.051)	(0.040)	(0.028)	(0.021)
Prev. Elected Exp.	0.007	−0.002	−0.014	0.030
	(0.047)	(0.039)	(0.026)	(0.021)
Incumbent	−0.015	0.051	0.007	0.029
	(0.040)	(0.034)	(0.022)	(0.018)
Non-Male	−0.145*	−0.133*	−0.044*	−0.041*
	(0.035)	(0.030)	(0.019)	(0.016)
Open	0.043	0.044	−0.011	0.014
	(0.043)	(0.036)	(0.023)	(0.019)
FE: Year-State	✓	✓	✓	✓
Constant	0.540	−0.555*	−1.161*	−1.010*
	(0.378)	(0.261)	(0.156)	(0.104)
Observations	1,663	2,529	1,663	2,529
R^2	0.585	0.585	0.890	0.890
Note:				$^*p<0.05$

References

Akaike, Hirotugu. 1974. "IEEE transactions on automatic control." *Econometrica* 19(6): 307–333.

Clarke, Kevin A. 2007. "A simple distribution-free test for nonnested model selection." *Political Analysis* 15(3): 347–363.

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

Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. 2022. *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton University Press.

King, Gary, Patrick Lam, and Margaret E Roberts. 2017. "Computer-assisted keyword and document set discovery from unstructured text." *American Journal of Political Science* 61(4): 971–988.

Lau, Jey Han, and Timothy Baldwin. 2016. "An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation." *1st Workshop on Representation Learning for NLP* .

Laver, Michael. 2017. "Measuring policy positions in political space." *Annual Review of Political Science* 17(1): 207–223.

Park, Ju Yeon, and Jacob M. Montgomery. 2023. "Towards a framework for creating trustworthy measures with supervised machine learning.".

Vuong, Quang H. 1989. "Likelihood ratio tests for model selection and non-nested hypotheses." *Econometrica* 57(2): 307–333.

Wilkerson, John, and Andreu Casas. 2017. "Large-scale computerized text analysis in political science: Opportunities and challenges." *Annual Review of Political Science* 20(1): 529–544.