Positioning in Congressional Primary Campaigns*

Mellissa Meisels[†]

May 14, 2023

Abstract

The concept of policy platform announcement is integral to theoretical investigation of elections, representation, and political behavior, yet empirical studies rely upon proxy measures that may not reflect candidates' public campaign positions. Leveraging original data on issue platforms from the campaign websites of 2016—2022 primary candidates, I introduce a new measure based directly on candidates' own campaign positions during the increasingly–important primary election stage. Aside from capturing candidates' strategic rhetoric unmediated by third-party actors, primary campaign positions are temporally dynamic, straightforward to validate, and clearly recover a recognizable liberal–conservative dimension. After extensive aggregate–, individual–, and word–level validation, the utility of the measure is demonstrated with an application to an ongoing debate regarding nationalization severing candidates' district ties. Estimating primary candidates' positions independently of campaign receipts and roll-call voting facilitates investigation into the substantive relationships between public campaign rhetoric, financial contributors, and subsequent legislative behavior.

[†]PhD Candidate, Department of Political Science, Vanderbilt University. `mellissa.b.meisels@vanderbilt.edu`.

Candidate positioning is one of the most ubiquitous components of both classic theories of elections and ongoing empirical debates in electoral and representation studies.[1] Announcement of policy platforms is an explicit focus in electoral behavior and competition in Downsian formal models (Banks 1990; Baron 1994; Cameron and Enelow 1992; Downs 1957; Enelow and Hinich 1982), and an implicit focus in scholarly debates regarding extremist success, the role of nationalization versus district preferences, and other normatively important topics within the study of legislative elections (Ansolabehere, Snyder, and Stewart 2001; Kujala 2020; Bonica and Cox 2018; Hall 2015; Hall and Snyder 2015; Hopkins 2018; Utych 2020; Woon 2018).

Understanding the dynamics of candidate positioning in modern primary elections is particularly important for two reasons. First, the primary stage has become increasingly important: the shrinking number of congressional districts competitive for both Democrats and Republicans today means that many House elections are as good as decided once the primary race ends (Abramowitz, Alexander, and Gunning 2006; Jacobson 1990, 2015).[2] Second, primaries are fought between those who share a party label, thus presenting an opportunity to select between different *kinds* of Republican or Democrats. The irrelevance of traditional partisan heuristics creates potential for a greater role of intra-party variation in positions, as candidates may be incentivized to distinguish themselves from a co-partisan field.[3]

Although the focus on primary elections has increased in tandem with their influence on shaping the contemporary Congress (e.g. Hassell 2023; Henderson et al. 2022; Hirano and Snyder 2019; Thomsen 2022), our ability to investigate key questions related to how

---

[1]While terms like *ideology* or *ideal point* are used frequently in studies involving collections of political views, this paper is interested solely in how candidates present themselves during elections and is apathetic about the "truthfulness" of these self-presentations. For this reason, I instead refer to *positioning* and *positions*, although I use descriptors such as *liberal*, *conservative*, and *extreme* that are commonly associated with ideology.

[2]The number of House races decided within 10% was 33 in 2016, 90 in 2018, 77 in 2020, and 75 in 2022 (`https://ballotpedia.org/Congressional_elections_decided_by_10_percent_or_less,_2018`).

[3]This is especially true in the case of open-seat primaries, where candidate fields tend to be large and no hopeful possesses the valence advantages enjoyed by a typical incumbent.

candidates position themselves publicly during primary campaigns is limited by existing data sources and measures. As proxies for candidate positioning, scholars typically rely on estimates of campaign contribution networks (Bonica 2013, 2014; Hall and Snyder 2015) or legislative roll-call voting (Clinton, Jackman, and Rivers 2004; Poole et al. 2011). While these measures possess appealing features and have been invaluable for advancing the study of electoral and legislative behavior, their assumptions and data sources make them unsuitable for use in certain settings. For instance, analyses of relationships between candidates' positions and either campaign contributions or legislative behavior require a measure of positioning estimated separately from the variable of interest on the other side of the equation. Additionally, questions related to how candidates publicly position themselves during a given election — whether for voters, co-partisans, potential donors, or the media — call for a measure that captures candidates' self-presentation.

I introduce a measure of positioning that is akin to the concept of policy platform announcement invoked by Downsian spatial models and sourced directly from primary candidates' own campaign rhetoric. Using original text data on issue positions collected from campaign websites, I scale election-specific, unidimensional estimates of primary candidates' positioning based on variation in word usage. This collection encompasses the over 6,000 candidates who appeared on major-party primary ballots in 2016, 2018, 2020, or 2022, allowing researchers to characterize the policy platforms and positioning of candidates from the most recent primary cycles.

The proposed measure of candidate positioning offers a number of conceptually and methodologically desirable properties. First, campaign websites capture candidates' issue priorities and positions in their own words — unmediated by media portrayals, donors' perceptions, or a fixed agenda (Druckman, Kifer, and Parkin 2009; Porter, Treul, and Mc-Donald N.d.) — yet this strategic rhetoric is likely influenced by a host of factors of scholarly interest, such as electoral and candidate characteristics. Additionally, the measure is transparent and straightforward to validate: candidates' individual estimates can be com-

pared to their plain-English campaign platforms, and word-level parameters recovered during scaling make clear how each word influences the position estimation. Finally, no assumptions are made about individuals' movement from one election to another, effectively producing a time series of positions for candidates who ran in multiple cycles over the period.

The new measure of campaign positioning is introduced as follows. First, I explain how campaign websites constitute an ideal source of data for capturing the concept of interest in a large swath of research on polarization and legislative elections. I then outline the process of collecting original data on issue positions from campaign websites and provide descriptive statistics on the primary candidates who are and are not captured in the data. Having shown the representativeness of those included, I introduce the text scaling model and algorithm used to estimate candidate-year positions based on word usage and frequency in campaign platforms. With estimates of candidate- and word-level parameters in hand, I probe the measure's construct and face validity, underlying dimensionality, and statistical relationship with external measures. Next, I apply the new measure to an ongoing debate regarding national versus constituency influence and demonstrate that existing measures would lead to different results. The concluding section explicates the broad utility (and limitations) of the measure for yielding new insights about congressional elections and how candidates' strategic self-presentations relate to their fundraising and future legislative behavior.

1 Measuring Candidate Positioning

Measures of political actors' positions are integral to many of the most important ongoing debates in every corner of our discipline. A proliferation of data and methodologies have advanced our ability to scale preferences for more and more groups of interest, yet the behavioral and statistical models underlying readily-available measures do not reflect

the concept that is often of interest in studies of representation, electoral behavior, and polarization: how candidates publicly position themselves during an election. This section explicates the gaps between this concept and existing approaches, the suitability of campaign websites as a source of positioning data, and the text scaling model used for position estimation.

1.1 Existing Approaches

The introduction of roll-call record-based ideological estimation transformed the study of individual-level legislative and electoral behavior. NOMINATE and its various flavors (McCarty, Poole, and Rosenthal 2006; Poole and Rosenthal 1991; Poole 2005), as well as Bayesian approaches that facilitate incorporation of external information (Clinton, Jackman, and Rivers 2004), allowed for the systematic characterization of congressmembers' spatial ideal points. While these methodologies opened the door for testing theories of representation and beyond (e.g. Bafumi and Herron 2010; Brady, Han, and Pope 2007; Canes-Wrone, Brady, and Cogan 2002), an increasing interest in the source of polarization called for a comparable measure of non-incumbent candidates' positions.

In response, Bonica (2014), Hall and Snyder (2015), and Hall (2015) leverage campaign receipts to generate estimates of candidates' positions based on their donation networks. These scalings rely crucially on the assumption that donors contribute expressively on the basis of ideological similarity to candidates, a behavioral model called into question by the apparent importance of strategic factors, such as district competitiveness and opponent ideology, to donors' decisions (e.g. Barber 2016; Barber, Canes-Wrone, and Thrower 2017; Gimpel, Lee, and Pearson-Merkowitz 2008; Meisels, Clinton, and Huber N.d.).[4]

Nevertheless, contribution-based estimates of candidates' positions (most notably, those from Bonica's (2014) Database on Ideology, Money, and Elections) offer an attractive option to alternative sources of data on legislative candidate behavior. Even major surveys of

---

[4]Although the same studies demonstrate that ideological congruence is one key factor in individual donors' decisions, widely-used contribution-based measures assume that it is the sole donation motivation.

federal candidates' stances have been plagued by low response rates for decades (McGhee et al. 2014), and television advertisements are prohibitively expensive for a large portion of House general election candidates, much less primary candidates (Herrnson, Panagopoulos, and Bailey 2020). And although researchers have derived text-scaling estimates of candidates' positions using Twitter data (Cowburn and Sältzer N.d.; Temporão et al. 2018), the inherently responsive nature of social media posts to events and controversies likely distinguishes Tweets from more stable collections of issue stances and policy platforms.

1.2 Why Campaign Websites?

Campaign websites constitute a uniquely well-suited source of data for estimating primary candidates' positioning. The vast majority of websites contain a page or section clearly delineated as a collection of issue stances, resembling the concept of a stated policy platform more closely than any other campaign activity. Moreover, the priorities and positions found on websites are selected and articulated by the candidate herself without the direct intervention of donors' perceptions, media portrayals, or constraining agenda items. Given that campaign websites "provide an unmediated, holistic, and representative portrait of messages aimed at voters in general" (Druckman, Kifer, and Parkin 2009, p. 346-347), scholars have long recognized their suitability for studying a variety of topics related to campaign strategy (Druckman, Kifer, and Parkin 2009; Druckman et al. 2010; Milita, Ryan, and Simas 2014; Nyhan and Montgomery 2015; McDonald, Porter, and Treul 2020; Porter, Treul, and McDonald N.d.).[5]

In addition to their maximally unfiltered and unconstrained picture of candidates' rhetoric, websites are also a highly accessible campaign medium compared to other activities. Creating and maintaining a website is trivially easy and cheaper[6] than fundraising,

---

[5]The scope of existing research using House campaign websites has been limited to general election candidates, or to primary candidates from one or two election cycles.

[6]While some candidates host highly professionalized websites clearly created by web designers, many candidates utilize free website creators, which offer easy-to-use interfaces that make website creation accessible to even the least technologically savvy candidate without the aid of campaign staff.

sending mailers, running television advertisements, and the vast majority of other potential sources of data on candidates' positions. Because the baseline costs involved in creating a website are so low, "missingness" in the data is more plausibly related to primary candidates' true decision not to publicly commit to a platform[7] than to factors orthogonal to a presence of positioning, such as insufficient fundraising (in the case of contribution-based measures) or failure to win election (in the case of roll-call based measures).

1.3 Data Collection: Primary Campaign Websites, 2016 — 2022

To characterize the rhetorical positioning of modern House primary candidates, I collect original data on the issue positions of all candidates who appeared on the ballot in a Democratic or Republican primary in 2016, 2018, 2020, and 2022 from campaign webpages.[8] This effort, which spans over 6,000 unique candidate-year observations, represents the largest collection of congressional primary candidates' platforms to my knowledge.
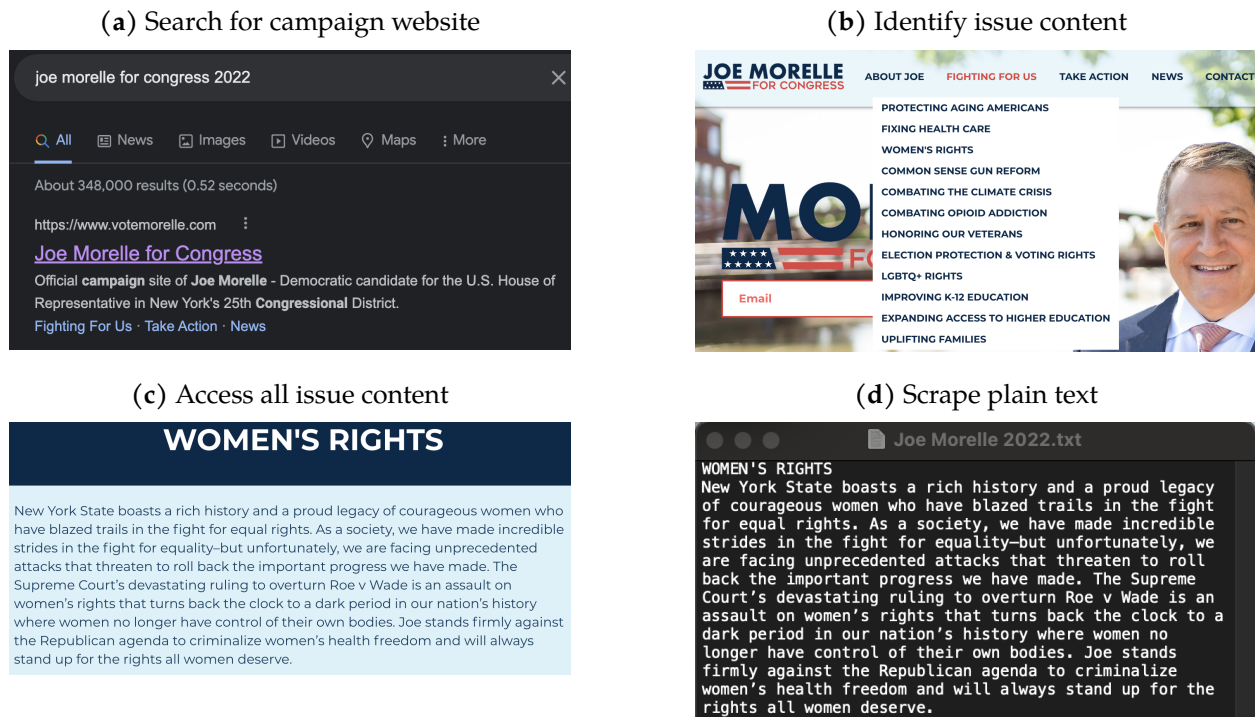
A simplified example workflow is illustrated in Figure 1, and Appendix A details all data collection steps at length. For each House district in a given election year, all candidates who appeared on the ballot in a Democratic or Republican primary were identified from *Ballotpedia*, which also provides basic information about the election and candidate. Next, I searched for each candidate's campaign website by Googling "[candidate name] for Congress [election year]" and cross-checking websites such as Politics1.com and the candidate's *Ballotpedia* page for a designated campaign website.[9] I used *Wayback Machine*

---

[7]As discussed shortly, while candidates who failed to raise non-trivial primary funds have a higher rate of campaign website position missingness, the rates are not substantially different from those with more resources or even incumbents. Considering the potentially large number of primary candidates who file FEC paperwork to run and then decide not to actively campaign, this suggests that campaign website missingness is related to a true lack of public positioning.

[8]Third-party candidates, candidates whose primaries were cancelled, and candidates in CA, WA, LA, CT, UT, and certain party primaries in some VA districts in some years are dropped. Appendix A provides the full list of and explanations for excluded locales.

[9]I exclude official governmental websites (those ending in .gov), as sitting incumbents maintain separate online presences for their campaign. Additionally, I exclude social media pages such as Facebook and Twitter, which are primarily forums for candidates to provide updates or respond to current events rather than establish stable platforms.

Figure 1. Example Data Collection Workflow

(**a**) Search for campaign website

(**b**) Identify issue content

(**c**) Access all issue content

(**d**) Scrape plain text

Note: Visual depiction of simplified steps involved in collecting Representative Joe Morelle's 2022 primary campaign issue positions from `www.votemorelle.com`. Appendix A describes each component of the data collection in detail.

to find the websites of candidates who ran prior to 2022 as archived most directly prior to the candidate's primary date, while 2022 candidates' websites were found in real time.[10]. I then navigated to issue content, which was typically found on a page or in a section clearly designated "Platform", "Issues", or "Priorities". Candidates' issue positions were manually scraped by copying and pasting the text into files as well as capturing an image of the content exactly as it appeared.

To investigate the representativeness of primary candidates who did and did not host campaign websites, Table 1 reports relationships between the binary presence of campaign website positions and observable candidate, election, and district characteristics thought to relate to candidates' willingness and ability to announce a platform. Data on fundraising are from FEC pre-primary reports and presidential vote shares are from Daily Kos,

---

[10]Candidates' live websites were accessed within a week of their primary election.

which include 2020 election results for post-census 2022 districts. I estimate models separately by incumbency status due to different meanings of missingness: virtually all incumbents hosted primary campaign websites over the period but some omitted positions, whereas the vast majority of missing non-incumbents lacked a campaign website altogether. Competition is captured by indicators for whether the primary was unopposed, financially uncompetitive, or financially competitive[11] as well as the party's advantage in the district[12] (Bartels 1986; Druckman et al. 2010; Lachat 2011; Grimmer 2013). In the non-incumbent model, I also indicate state legislative experience[13] and whether a candidate raised under 10% of the total receipts in the primary (Milita, Ryan, and Simas 2014).

Table 1 indicates high rates of campaign website position-taking, especially (and unsurprisingly) among incumbents and those who garnered more than a diminimous share of their primary's total fundraising. Non-incumbents who raised under 10% of the total receipts are 15% less likely to have website positions than those who raised over. However, the magnitude of this missingness is relatively modest considering that nearly 40% of non-incumbents did not even file pre-primary fundraising reports, and a substantial portion of such candidates likely did not actively campaign after filing to run. Overall, the results do not suggest that large swaths of candidates are systematically excluded from data on campaign website positions on the basis of candidate type, electoral competitiveness, or even resources.

1.4 Scaling Primary Campaign Positions

Having provided evidence that those who take positions are broadly representative of the universe of primary candidates, I now turn to estimating candidates' overall primary po-

---

[11] Following Thomsen (2022), a primary is coded as financially competitive if the top fundraiser garnered under 57.5% of the total receipts in the primary.

[12] Following Hirano and Snyder (2019), a party is advantaged if their nominee received over 57.5% of the vote share in the most recent presidential election, disadvantaged if they received under 42.5%, and swing if their vote share was somewhere in between.

[13] I thank Alan Wiseman for sharing data on state legislators.

Table 1. Determinants of Primary Campaign Website Positions, 2016—2022

	Campaign Website Positions Present	
	Incumbents	Non-Incumbents
(Intercept)	0.866***	0.775***
	(0.068)	(0.028)
Republican	0.025	−0.014
	(0.027)	(0.014)
Unopposed Primary	−0.087	−0.054
	(0.065)	(0.028)
Uncompetitive $ Primary	−0.050	−0.014
	(0.065)	(0.017)
Advantaged District	−0.076**	
	(0.028)	
Receipts $< 10\%$		−0.146***
		(0.015)
State Legislator		0.025
		(0.026)
Open Advantaged		0.001
		(0.025)
Open Disadvantaged		−0.068*
		(0.034)
General Challenger Swing		−0.009
		(0.024)
General Challenger Disadvantaged		−0.092***
		(0.023)
Primary Challenger Advantaged		−0.046
		(0.025)
Primary Challenger Swing		−0.029
		(0.033)
Year Fixed Effects	✓	✓
Observations	1,213	4,939
Adjusted R^2	0.012	0.100

Note: Linear probability models predicting presence (1) or absence (0) of campaign website issue positions during primary. Reference value for primary competitiveness is financially competitive, district type in incumbent model is swing, and district-candidate type in non-incumbent model is open-seat swing. HC3 standard errors in parentheses. *$p<0.05$; **$p<0.01$; ***$p<0.001$

sitioning based on their campaign website issue text. I follow other scholars in assuming that the frequency and usage of words in political text are informative about authors' positions on a liberal–conservative dimension (Lauderdale and Herzog 2016; Laver, Benoit, and Garry 2003; Lowe et al. 2011; Rheault and Cochrane 2020; Vafa, Naidu, and Blei 2020).

As demonstrated by Grimmer and Stewart (2013), however, the validity of this assumption rests crucially on the dominance of a liberal-conservative dimension within the relevant texts. Manually identifying issue positions ensures that the collection of campaign website text is laser-focused on issue positioning content, and the proceeding section provides individual–, aggregate–, and term–level evidence to extensively validate the underlying dimensionality structuring primary campaign discourse.

To prepare the campaign position text corpus, I construct an $\mathbf{M} \times \mathbf{N}$ sparse document-feature matrix with $\mathbf{M}$ term columns, $\mathbf{N}$ candidate-year rows, and term frequencies as cell entries. I pre-process the data by removing punctuation, reducing terms to their stem, and removing both highly frequent stopwords and highly infrequent terms to reduce noise in estimation and improve computing time.[14] To help ensure that the key liberal–conservative dimension is identified and minimize the risk of misspecifying the dimensionality of the policy space, I drop terms primarily associated with geographic or incumbency differences between candidates, such as state names and congressional procedure. In addition to all remaining unigrams that meet the above criteria, I likewise preserve a number of frequently-used bigrams (e.g. `common core`), trigrams (e.g. `freedom of speech`), and quadgrams (e.g. `right to bear arms`).[15] Altogether, this results in more than 2,500 unique terms across over 3,800 separate primary campaigns. The detailed text processing flow and comparisons of subsequent estimates with and without scaling refinements are provided in Appendix B.

I use an unsupervised machine learning algorithm, `wordfish`, to scale unidimensional primary campaign positions at the candidate-year level (Slapin and Proksch 2008). The statistical model is based on item response theory and bears strong resemblance to correspondence analysis, the methodology used to estimate campaign contribution-based CF

---

[14]I drop terms that appear in fewer than 100 separate campaign texts. This is an extremely lenient requirement given that the corpus contains almost 4,000 campaign texts, yet this step substantially improves computing time. See Appendix B for further discussion of pre-processing choices.

[15]Scatterplots in Appendix B demonstrate tight relationships between scaling with and without non-unigram, geographic, and procedural terms ($r = 0.997$; $\rho = 0.998$).

Scores (Bonica 2014).[16] Importantly, the model also accounts for candidate-level differences in wordiness and word-level differences in the informativeness *vis-à-vis* candidates' positions.[17] The rate y at which primary candidate i uses term j in election year t is assumed to be drawn from a Poisson distribution, which is characterized by a single parameter λ representing both the expected value and variance. This parameter logarithmically links the probability distribution generating the observed term rate to the systematic components of interest:

$$y_{ijt} \sim \text{Poisson}(\lambda_{ijt}) \text{ where } \lambda_{ijt} = \exp(\alpha_{it} + \psi_j + \beta_j * \omega_{it}). \tag{1}$$

The key parameter is ω, which stands in for candidate i's latent primary campaign position in election t, and is scaled to have sample mean 0 and standard deviation 1. β represents word j's weight in discriminating between different campaign positions.[18] A word fixed effect ψ captures the rate at which word j is used generally, and a candidate-year fixed effect α corresponds to the verbosity of candidate i's campaign position text in election t. After calculating start values,[19] estimation proceeds via expectation maximization, which entails estimating conditional expectations for the word and candidate paramaters, calculating conditional maximum likelihoods, and iterating using these new parameter expectations until the model converges successfully. Further technical details of the text data pre-processing, algorithm initialization, and parameter estimation, as well as alternative scalings using correspondence analysis and unrefined tokens, are relegated to Appendix B.

---

[16]Scatterplots in Appendix B demonstrate tight relationships between scaling estimates from `wordfish` and one-dimensional correspondence analysis ($r = 0.988$; $\rho = 0.998$).

[17]For example, the term `gun` is neutral and used by candidates all across the political spectrum, whereas the term `high-capacity` highlights the danger of large firearm magazines and thus predominantly associated with candidates on the left.

[18]This is akin to an IRT discrimination parameter or factor analysis loading score.

[19]Start values of ψ and α are functions of word frequencies, while start values of β and ω are obtained via singular value decomposition of the matrix of word frequency marginals — hence the strong relationship between estimates resulting from correspondence analysis versus `wordfish` in the Appendix.

2 Validation and Comparisons to Existing Measures

With estimates of primary campaign positioning — the ω parameters from Equation 1 — in hand, I subject the new measure to a series of validation exercises and consider its relationship to alternative measurements. First, I show that the measure replicates the well–known bimodal distribution that distinguishes between positions of Republicans and Democrats. Next, I establish the face validity of selected candidates' campaign position estimates across the range of the measure. I then probe the dimensionality of the measure by examining the underlying words that structure campaign discourse and endogenizing the scaling by performing year– and incumbency–specific estimation. Finally, I explore relationships between primary campaign positions, DIME, and DW-NOMINATE.

2.1 Aggregate Distribution

Estimates of primary campaign positions by candidate partisanship are presented in Figure 2. Consistent with well–documented partisan polarization among political elites (Bafumi and Herron 2010; McCarty, Poole, and Rosenthal 2016; Theriault 2006; Thomsen 2014), positions are bimodally distributed, with most Republicans substantially to the right of most Democrats and most Democrats substantially to the left of most Republicans. However, a modest degree of overlap in Republican and Democratic candidates' positions is likewise consistent with the frequency with which candidates of both parties choose to campaign similarly on the same issues, such as job creation and veterans affairs. This contrasts with roll-call estimates of House members' ideal points from recent congresses, which exhibit no partisan overlap partly due to the strategic selection of legislative floor votes that frequently exaggerate differences between partisan voting coalitions and their members' roll call-based positions (Clinton 2012; Clinton and Lapinski 2008; Cox and McCubbins 2005; Lee 2016).

Figure 2. Distribution of Primary Campaign Positions by Party, 2016 – 2022
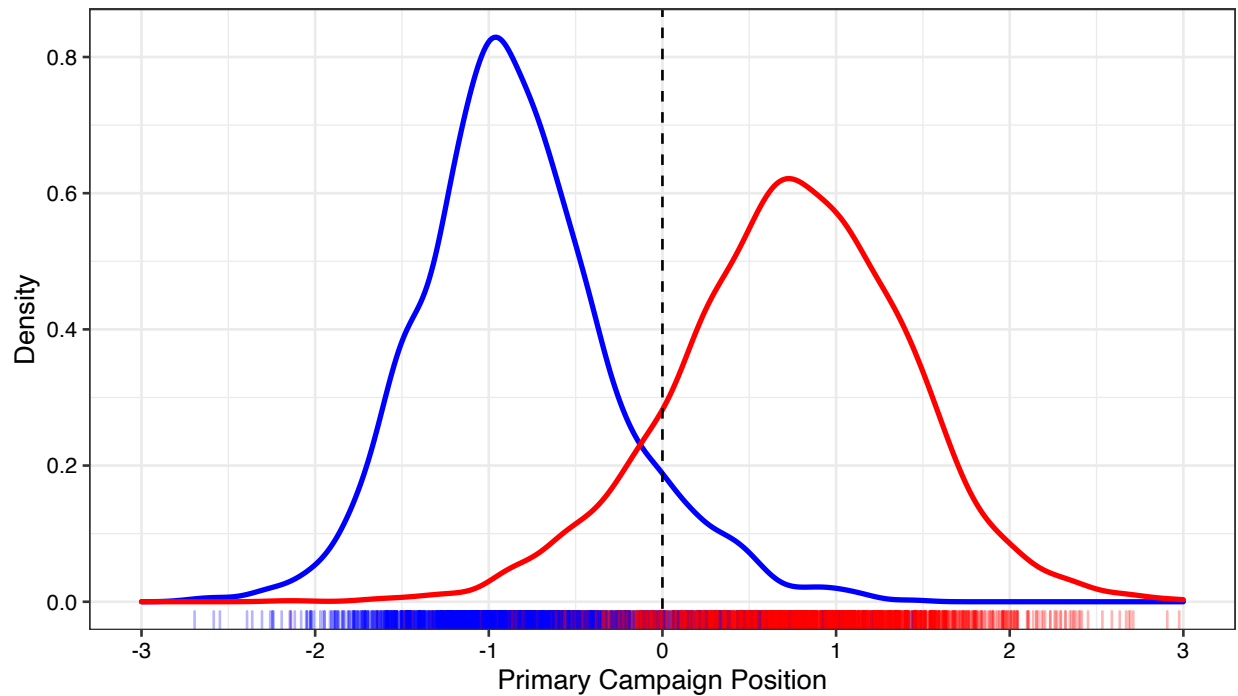
Note: Kernel density plots of ω estimates from Equation 1, representing primary candidate positions based on campaign website platforms. Democratic candidates in blue and Republican candidates in red. Negative values indicate more liberal/less conservative.

2.2 Selected Candidates' Primary Campaign Positions

In addition to the measure's ability to distinguish Republicans from Democrats, we can also assess the face validity of candidates' primary campaign positions to evaluate whether the observed intra-party variation appears to comport with reality. A selection of relatively well–known candidates' positions from across the political spectrum is reported in Table 2. A Republican representing a rural district in New York's North Country since 2014, Elise Stefanik's 2018 primary campaign fell almost one standard deviation to the left of the mean. Although she has since made headlines for her impassioned defense of former President Trump during his first impeachment proceedings[20], Stefanik campaigned on strengthening trade with Canada, expanding agricultural visa programs, veteran wel-

---

[20]https://www.reuters.com/world/us/loyalty-trump-catapults-elise-stefanik-into-republican
-stardom-2021-05-11/

Table 2. Primary Campaign Positions of Notable Candidates

Candidate	District-Year	Campaign Position
Elise Stefanik (R)	NY-21-2018	−0.93
George Devolder-Santos (R)	NY-3-2020	−0.13
Liz Cheney (R)	WY-2022	0.29
Andy Biggs (R)	AZ-5-2022	0.58
John Rose (R)	TN-6-2022	1.07
Madison Cawthorn (R)	NC-11-2022	1.35
Marjorie Taylor Greene (R)	GA-14-2020	1.95
Henry Cuellar (D)	TX-28-2016	0.21
Jim Cooper (D)	TN-5-2020	−0.00
Debbie Wasserman Schultz (D)	FL-23-2020	−0.21
Joaquin Castro (D)	TX-20-2018	−0.74
Jerrold Nadler (D)	NY-12-2022	−1.01
Alexandria Ocasio-Cortez (D)	NY-14-2022	−1.54
Ayanna Pressley (D)	MA-7-2020	−1.89

Note: Primary campaign positions of selected candidates from across the scale. Campaign positions are ω estimates from Equation 1, which are scaled to have mean 0 and standard deviation 1.

fare, environmental protection, healthcare access, and affordable education. Conversely, the 2016 primary of Texas Democrat Henry Cuellar, who has voted with Republicans on legislation regarding abortion, firearms, and immigration[21], was almost a quarter standard deviation to the right of the mean. Tennessean Blue Dog Democrat Jim Cooper, the "man in the middle"[22] and "the last moderate...loathed by Republicans for being in the wrong party, and scorned by Democrats for his fiscal conservatism"[23] represented mean 0 during his 2020 primary campaign. Likewise, the campaigns of those widely regarded as the most progressive Democrats and conservative Republicans fall to the extremes of the campaign position range.

---

[21]https://www.texastribune.org/2022/10/06/henry-cuellar-texas-2022/
[22]https://washingtonmonthly.com/2022/12/13/man-in-the-middle/
[23]https://www.nytimes.com/2011/09/06/opinion/the-last-moderate.html

2.3 Content and Dimensionality

Of chief importance for validating unsupervised scaling methods is ensuring that the dimension of interest — here, a left-right, issue-based dimension — is the one structuring individuals' positioning estimates (Grimmer and Stewart 2013; Egerod and Klemmensen 2020). Luckily, interrogating the underlying substance is relatively straightforward and transparent in the case of text data, as words included in the scaling likewise receive parameter estimates based upon their ability to discriminate between positions. Table 3 reports the ten terms with the largest negative (left) and positive (right) β weights from Equation 1, while the top 20 terms and their corresponding β and ψ estimates are reported in Appendix B. While words related to critical race theory, Christianity, anti-abortionism, illegal immigration, and socialism are strongly associated with conservative campaign positions, terms related to inequality, injustice, gender and sexuality, and affordable education are strongly associated with liberal positions. Overall, these results provide strong evidence that the rhetoric underlying the scaling estimates is structured by well–recognized modern divisions along the liberal–conservative spectrum.

A major advantage of primary campaign positions is the dynamic, time-series nature of the measure: if a candidate's campaign rhetoric changes from primary-to-primary, so too will her estimated primary campaign position. While the narrow temporal and Trump-era scope of the data makes it especially unlikely that the meaning of words changed substantially across the time period (Egerod and Klemmensen 2020), it is nevertheless informative to check whether the vocabulary of primary campaigns differed from one election to the next. Performing scaling separately by year in Appendix B suggests substantial continuity in even the top terms with the most liberal and conservative weights,[24] as well as correlations above 0.90 between primary campaign position estimates from the pooled scaling

---

[24]Moreover, the emergence of heavily weighted terms such as `lewi` (a stem from references to the John Lewis Voting Rights Act, legislation proposed by House Democrats in the 117th congress) and `crt` in 2022 is consistent with contemporaneous real-world changes in Democrats' and Republicans' electoral and legislative priorities.

Table 3. Words With 10 Most Conservative and Liberal Weights

Left	Right
community-bas, rental, equit, reproduct, trauma, matern, lgbtq, high-capac, lewi, low-incom, dispar, childcar, disproportion, pell, pre-k, tuition-fre, underserv, expung, resili, discriminatori	critical race theori, build the wal, tyrann, crt, indoctrin, god, tyranni, christian, sanctiti, god-given, unborn, pro-lif, communist, swamp, socialist, amnesti, islam, sanctuary c, alien, 2nd amend

Note: Terms with the ten largest positive (right) and negative (left) β discrimination parameters from scaling. Appendix B reports top 20 terms and corresponding β and ψ parameters.

and each of the year-specific scalings.

An additional benefit of campaign websites as a data source for candidate positions is the inclusion of vastly understudied long-shot candidates. Although little can be said about the financial contributors, campaign advertisements, or expenditures of candidates who did not file pre-primary fundraising reports with the FEC, 45% of such candidates nevertheless hosted campaign websites with positions, and are therefore included in the new measure of campaign positioning. However, to ensure that the scaling space is not primarily defined by marginal candidates who may be using rhetoric distinct from that of viable candidates, I perform the scaling including only sitting incumbents. The strong correlation of over 0.95 between incumbents' campaign positions from the pooled and incumbent-only scalings shown in Appendix B suggests that marginal candidates do not differentially drive nor distort campaign positioning estimates.

2.4 Comparisons to Existing Measures

As discussed previously, the most frequently used off-the-shelf proxies for candidate positioning are legislative voting-based DW-NOMINATE (Lewis et al. 2023) and campaign contribution-based CF Scores (Bonica 2019). The measure of primary campaign positioning introduced here differs conceptually and methodologically from these existing estimates in a number of important ways. While DW-NOMINATE and CF Scores place con-

straints on individuals' movement across time,[25] no special assumptions are made about the continuity of primary candidates' campaign positions from election-to-election.

Moreover, the new measure of positioning is based on candidates' own rhetoric and issue statements rather than sourced from behavior directly mediated by outside actors. Although candidates may strategically tailor their issue attention and positions to best appeal to potential donors, voters, and activists, candidates themselves ultimately decide on their campaign platforms. In contrast, the legislative agenda and especially roll-calls are selected strategically with little input from most members of Congress (Clinton 2012; Clinton and Meirowitz 2001; Lee 2016; Patty and Penn 2019). Similarly, donors do not contribute solely on the basis of candidates' positions (Barber, Canes-Wrone, and Thrower 2017; Meisels, Clinton, and Huber N.d.; Stuckatz 2022), and may form their own perceptions of candidates' positions based upon both direct and non-direct sources of information (Bonica 2014).

Although there is not necessarily reason to expect strong correlations, it is nevertheless descriptively useful to investigate the relationships between candidates' primary campaign positions and summaries of their roll call behavior and contribution networks. In Figure 3, I present scatterplots comparing the new positioning measure to CF Scores and DW-NOMINATE. The left plot includes primary candidates from 2016 and 2018 as CF Scores are only available through 2018, and the right plot includes only sitting legislators.[26]

The patterns in Figure 3 suggest that primary campaign positioning co-varies more strongly with future roll-call voting than with campaign contribution networks. While pooled overall correlations between the measures are quite strong (but still stronger with

---

[25]While more flexible varieties of NOMINATE have been developed (Lewis and Sonnet N.d.; Nokken and Poole 2004), the original NOMINATE only allowed for linear movement across time for legislators serving over 10 years (Poole and Rosenthal 1991; Poole 2005), and NOMINATE estimates available on `voteview.com` fix legislators across time (Lewis et al. 2023). CF Scores assume that recipients are fixed across time — including while running for different levels of office (Bonica 2014).

[26]Candidates who successfully won their election were matched to their DW-NOMINATE score from the following congress: the 2016 election corresponds to the 115th, 2018 to the 116th, 2020 to the 117th, and 2022 to the 118th.

17

Figure 3. Relationship Between Primary Campaign Scores, CF Scores, and DW-NOMINATE
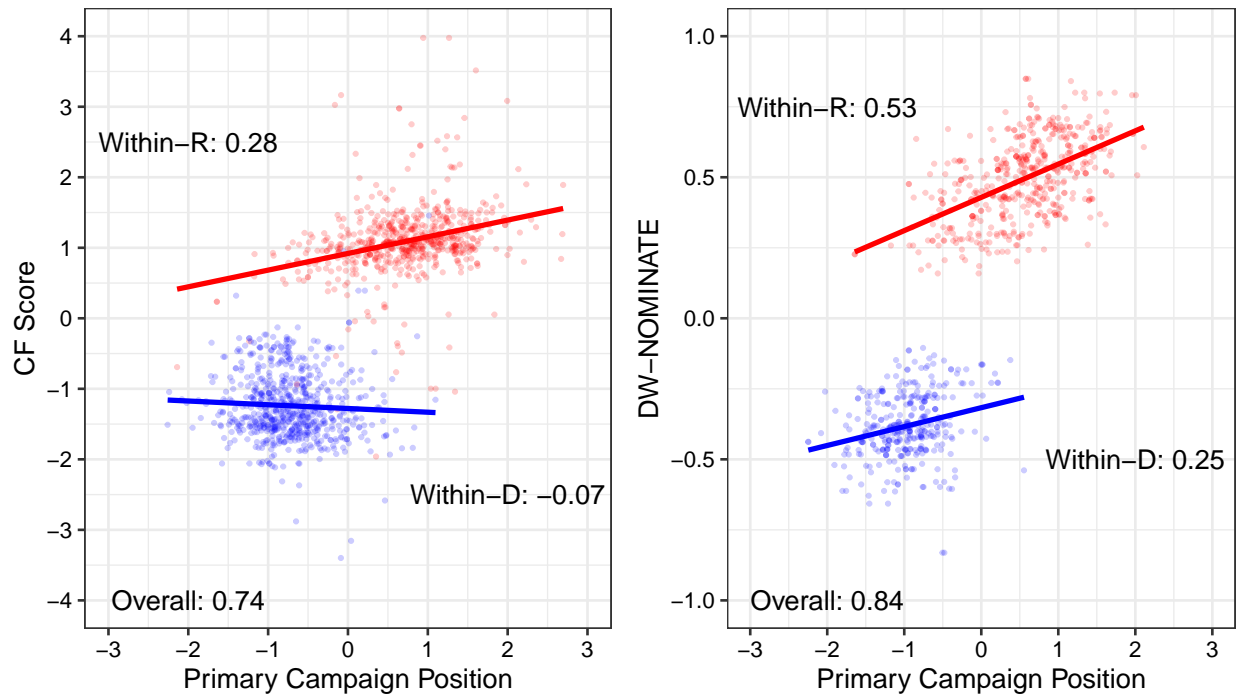
Note: Left plot compares the primary campaign positions and CF Scores of House candidates from 2016 and 2018. Right plot compares the primary campaign positions and 1st–dimension DW-NOMINATE scores of members of the 115th–118th Houses. Simple bivariate regression lines fit separately by party. Pooled and intra-party Pearson's r correlations reported.

DW-NOMINATE than with CF Scores), the intra-party correlations between primary campaign positions and DW-NOMINATE are substantially greater than those with CF Scores. Whereas recent work by Barber (2022) documents the complete disappearance of a statistical relationship between House Democrats' CF Scores and DW-NOMINATE scores since 2014, Democrats' primary campaign positions exhibit a persistent relationship with their DW-NOMINATE scores, and the NOMINATE–campaign position correlations for Republicans and candidates overall are comparable to the NOMINATE–CF Score correlations found in Barber (2022).[27] These results highlight the potential for further investigation of

---

[27]Moreover, the differences in strength of Democrats' relationships between the new measure and existing measures are not driven by the differential inclusion of non-incumbents when making comparisons to CF Scores versus DW-NOMINATE scores. The intra-Democrats correlation between campaign positions and CF Scores among only those with DW-NOMINATE scores remains at a paltry 0.03.

relationships between candidates' primary positioning, fundraising activity, and legislative behavior facilitated by measuring positions independently of campaign contribution and roll-call data.

3   Evaluating District Importance to Candidate Positions

Thus far, I have introduced and extensively validated a measure of campaign positioning based directly on House primary candidates' own public, election-specific rhetoric. I now turn to an example of the measure's utility for providing new insights into candidate behavior with a descriptive application to the ongoing debate regarding nationalization versus district preferences. Crucially, I demonstrate that relying instead upon contribution–based estimates would lead to a conclusion opposite of that reached using primary campaign positions.

3.1   Does the District Still Matter?

The importance of (sub-)constituency is all but a given in classic theoretical and empirical studies of elections and representation (e.g. Canes-Wrone, Brady, and Cogan 2002; Downs 1957; Enelow and Hinich 1984; Meirowitz 2005; Miller and Stokes 1963). Whether represented by the median or a distribution, and consisting of voters, constituents, co-partisans, or donors, the key population of interest in candidates' strategic positioning is thought to be district–specific. However, recent evidence on the nationalization of political behavior, media, and donors calls into question whether candidate–district ties have been severed (Ansolabehere, Snyder, and Stewart 2001; Abramowitz and Webster 2016; Gimpel, Lee, and Pearson-Merkowitz 2008; Hopkins 2018; Jacobson 2015; Martin and McCrain 2019; Moskowitz 2021).

Consistent with this, Bonica and Cox (2018) argue that political parties strategically nationalized congressional elections in response to increased competition for majority con-

trol since 1994 (Lee 2016). If elections are primarily fought over national party positions, national donor support, and national media attention, candidates no longer stand to benefit from tailoring their positions to the district, and instead stand to benefit from adopting the party line and appealing to extreme donors and activists.[28] However, more recent work finds that support for extreme nominees has not increased post-1994, suggesting that incentives have not changed along these lines (Lockhart and Hill N.d.).

Given that primary elections have become increasingly consequential, as the number of House districts competitive for both parties in the general election is in decline, I provide new evidence from the primary stage regarding whether candidates' positions vary systematically by district or nationalization has severed such ties. Focusing on primary candidates as the sample population presents a potentially more difficult case: the preferences of candidates' key primary constituency (whether co-partisan constituents, voters, or donors) are unlikely to perfectly co-vary with district preferences, which may induce an even weaker district–candidate relationship than would be found in the general election case.

3.2   Evaluating District–Candidate Ties

I investigate primary candidates' responsiveness to district preferences using by estimating the statistical relationship between district partisanship and both primary campaign positions and CF Scores. While campaign positions reflect primary candidates' public rhetoric, CF Scores represent the political leanings of candidates' individual donor network. As a first step, Figure 4 plots the candidate-year level bivariate relationships between district two-party Democratic presidential vote share and primary campaign positions, while Figure 5 plots district Democratic vote and CF Scores. Although CF Scores only include candidates from two primary cycles while primary campaign positions in-

---

[28]Specifically, Bonica and Cox (2018) argue that voters have become more party-centered and therefore no longer penalize candidates for extremism, whereas extremism can benefit candidates via activist and donor support.

clude candidates from four, plotting only primary candidates who also have CF Scores in Appendix C reproduces the trends shown in Figure 4.

Figure 4 shows that as the Democratic lean of districts grows stronger, both Democratic and Republican primary candidates' positions consistently become more liberal (or less conservative).[29] Although there is an unsurprisingly persistent intercept shift between candidates of opposing parties running in similar districts, the lines fit separately by party demonstrate strong relationships between primary candidates' positions and their district's lean that cannot be explained by candidates' partisan differences. On the other hand, Figure 5 paints a different picture in the case of candidates' contribution networks. A very flat red line suggests that Republicans' CF Scores do not become more conservative as their districts grow less Democratic, while Democrats' CF Scores appear to become *less* liberal in more heavily Democratic districts. Appendix X demonstrates that disparities between campaign positions and CF Scores trends are not due to sample differences.

To evaluate the magnitude of the descriptive relationships between primary candidates' positions and district preferences from 2016 to 2022, I estimate the following equation separately for Democrats and Republicans:

$$\text{Position}_{idt} = \alpha + \tau \text{District}_{dt} + \upsilon \text{Open}_{idt} + \kappa \text{GenChall}_{idt} + \eta \text{PrimChall}_{idt} + \gamma_t + \epsilon_{idt} \quad (2)$$

where Position stands in for two dependent variables scaled to have mean 0 and standard deviation 1: candidate i's campaign position ω from Equation 1 during the primary in district d in year t, and her recipient CF Score.[30] The key independent variable, District, represents district d's Democratic two-party vote share centered at 50% from the presidential election held in or most immediately before year t. Because primary campaign positions and CF Scores have a standard deviation of 1, multiplying parameter υ by 100

---

[29]The substantial spread of individual points makes it difficult to visually evaluate the linearity of the relationships. In Appendix C, fitting Loess curves instead suggests that the relationships are highly linear.

[30]Notation is abused slightly by indexing Position by ip, as CF Scores vary only at the candidate level.

Figure 4. Primary Candidates' Positions and District Partisanship, 2016–2022
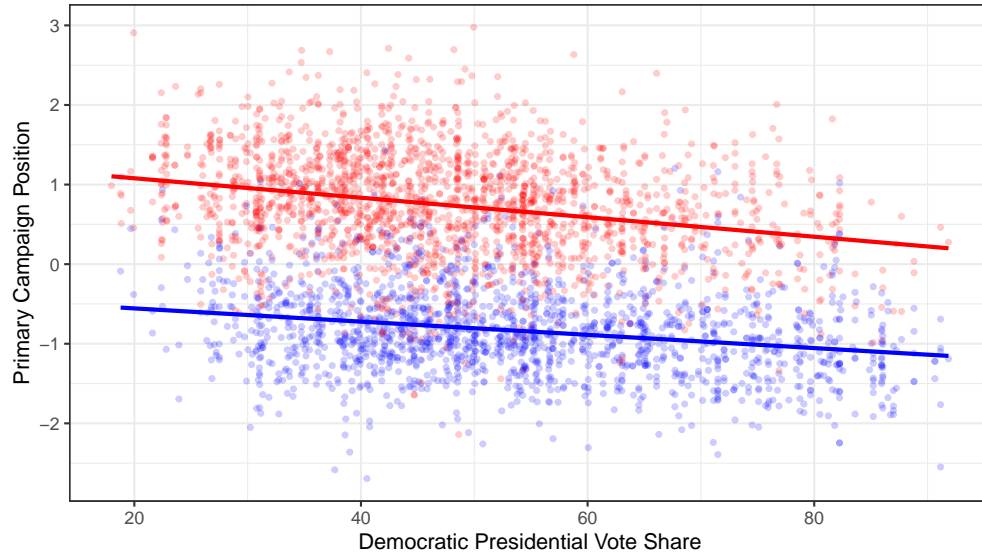
Note: Scatterplot points are individual candidate-year observations, with Republicans represented in red and Democrats represented in blue. Simple bivariate regression lines fit separately by party. Vertical axis plots ω from Equation 1, which increase with conservatism and are scaled to have mean 0, std. dev. 1. Horizontal axis plots district's most recent two-party Democratic presidential vote share.

Figure 5. CF Scores and District Partisanship, 2016–2018
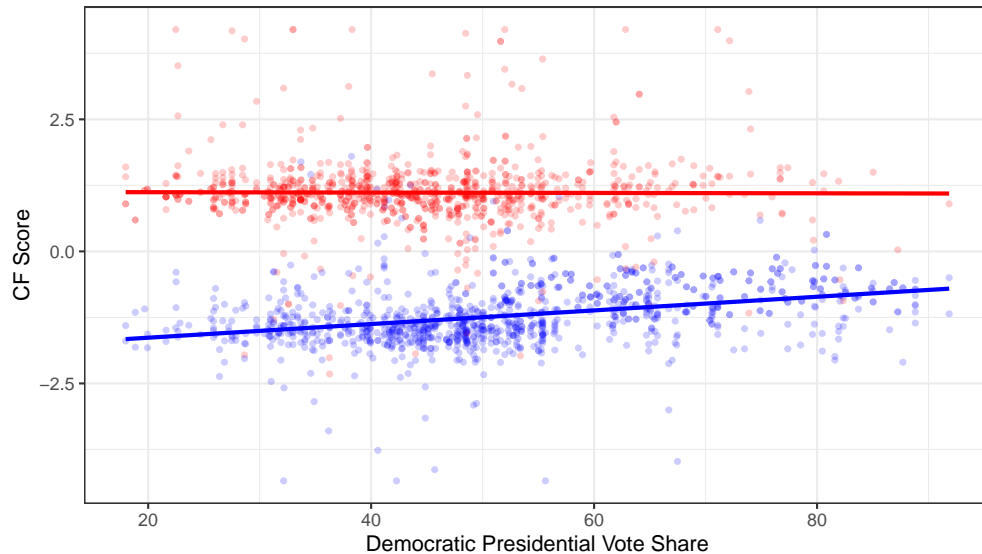
Note: Scatterplot points are individual candidate-year observations, with Republicans represented in red and Democrats represented in blue. Simple bivariate regression lines fit separately by party. Vertical axis plots recipient CF Scores, which increase with conservatism and are scaled in Bonica (2014) to have mean 0, std. dev. 1. Horizontal axis plots district's most recent two-party Democratic presidential vote share.

corresponds to the percentage of a standard deviation change in the outcome variable associated with increasing district Democratic vote by 1%. To examine descriptive differences between campaign positions by candidate type, indicator variables capture whether i was an open-seat candidate, a primary challenger, or a prospective general election challenger in primary dt. As such, intercept α represents the primary campaign position of an incumbent representing a district with equal Democratic and Republican presidential vote share. Finally, I include year fixed effects to account for secular trends in candidates' extremism, progressivism, or campaign issue focus and use HC3 standard errors.[31]

3.3   Conclusions Diverge By Measure

Table 4 reports the relationships between primary candidates' positions, the district's Democratic lean, and candidate type by positioning measure and candidate partisanship, and results from models using alternative district ideology estimates (Tausanovitch and Warshaw 2013; Warshaw and Tausanovitch 2022) as the key predictor or allowing district partisanship to interact with candidate type are presented in Appendix C. Among both Democrats and Republicans, primary campaign positions become significantly more liberal (or less conservative) as the district grows more heavily Democratic: increasing a district's previous Democratic presidential vote share by 10% is associated with Democratic primary candidates' campaign positions becoming 10% of a standard deviation more liberal, while Republicans' grow 14% of a standard deviation more liberal. This substantively meaningful result suggests that even during the primary, the Republican and Democratic pools of publicly-espoused campaign positions vary systematically by the district's partisan composition.

In contrast, campaign contribution networks do not appear to exhibit a similar relationship to district partisanship. While the coefficient corresponding to district Democratic lean achieves conventional levels of significance in the Democratic candidate model, it is

---

[31]Clustering standard errors at the primary level (e.g. the 2020 Republican primary in AL-1) is unsurprisingly immaterial to the results given the large number of primaries that are unopposed.

Table 4. Relationship Between District Partisanship and Candidate Positions

	Primary Campaign Position		Recipient CF Score	
	Democrats	Republicans	Democrats	Republicans
(Intercept)	−0.567***	0.489***	−0.733***	0.985***
	(0.036)	(0.038)	(0.041)	(0.030)
District Dem. Partisanship	−0.009***	−0.014***	0.003*	−0.004
	(0.001)	(0.001)	(0.001)	(0.002)
Open Seat Candidate	0.001	0.364***	−0.491***	0.183***
	(0.036)	(0.041)	(0.048)	(0.046)
Primary Challenger	0.203***	0.457***	−0.393***	0.259**
	(0.044)	(0.042)	(0.075)	(0.080)
General Challenger	0.018	0.347***	−0.586***	0.282***
	(0.039)	(0.049)	(0.045)	(0.065)
Year Fixed Effects	✓	✓	✓	✓
Observations	1,778	2,025	1,117	976
Adjusted R^2	0.135	0.159	0.208	0.025

Note: Parameters from Equation 2 with HC3 standard errors in parentheses. Predictors are district Democratic two-party vote from most recent presidential election and candidate type, with intercept representing an incumbent in a 50% Democratic district. Models 1 and 2 include 2016–2022 primary candidates with primary campaign positions ω from Equation 1. Models 3 and 4 include 2016 and 2018 primary candidates with recipient CF Scores from Bonica (2019). *p<0.05; **p<0.01; ***p<0.001

relatively small and signed in the unexpected direction: a 10% increase in Democratic presidential vote share in the district is associated with Democratic primary candidates having 3% of a standard deviation more *conservative* CF Scores. District partisanship is correctly signed in the case of Republican primary candidates, however, the relationship with CF Scores is similarly small and fails to reach statistical significance.

Additionally, Table 4 uncovers evidence that non-incumbent Republican primary candidates' contribution networks and campaign positions are both substantially more conservative than those of incumbent Republicans, but disparate trends emerge among Democratic primary candidates. Although Democratic open-seat candidates, primary challengers, and prospective general election challengers have far more liberal CF Scores than Democratic incumbents on average, the primary campaign positions of Democratic open-seat candidates and general challengers do not appear significantly more liberal than those of Democratic incumbents, and Democratic primary challengers have, on average, 20% of a

standard deviation *less* liberal campaign positions than Democratic incumbents. The extremism of non-incumbents relative to incumbents among Republicans and not Democrats is consistent with recent evidence regarding patterns of state legislators running for the House (Phillips, Snyder, and Hall N.d.).

While pinpointing the mechanism driving the relationship between district preferences and primary candidates' positions is beyond the scope of this brief application of the new measure, a number of potential explanations are consistent with the results presented here. First, candidate behavior may remain district–tailored while donor behavior has nationalized: the district–CF Score relationships in Figure 5 suggest donors may contribute to co-partisans across the nation, and thus along the political spectrum. Moreover, variation in primary candidates' public campaign rhetoric by district preferences can be explained by strategic candidate entry, strategic campaigning behavior, or simple differences in positions of potential candidate pools across districts.

4   Conclusion

The concept of a policy platform announcement is integral to theoretical investigation of elections, representation, and political behavior, yet empirical studies rely upon proxy measures that may or may not be related to candidates' public campaign positions. Using data collected directly from campaign website issue platforms, I introduce a new measure based on candidates' own campaign rhetoric during the increasingly–important primary election stage. By subjecting it to extensive validation, I demonstrate that the scaling successfully captures a widely-recognizable liberal-conservative dimension and derives facially valid estimates of primary candidates' campaign positions. Moreover, a simple application of the new measure to an ongoing debate regarding the severing of candidates from their districts highlights its utility and ability to provide unique insights that would be missed by relying on existing measures.

Every measure exhibits limitations, and this one is both no exception and should not serve as a universal substitution for alternative widely-used measures. Primary campaign positions do not purport to measure candidates' "true" ideologies, whether defined as sincerely-held beliefs or predictors of future behavior. As such, they should be employed in analyses concerned with candidates' public-facing priorities and issue stances. Additionally, the general adoption of campaign webpages in the 2000s at the earliest (Druckman, Kifer, and Parkin 2007; Sulkin, Moriarty, and Hefner 2007) and the time– and manpower–intensive nature of systematically collecting positions from primary candidates' websites makes it infeasible to match the quantity of estimates on par with measures based on more readily-available data: primary campaign positions currently cover four election cycles (albeit the most very recent ones), while CF Scores extend from 2018 back through the 20th century, and DW-NOMINATE offers coverage of every U.S. Congress. Moreover, a lack of comparable sources of issue platforms from members of the public and other non-political actors precludes a common-space scaling.

Aside from temporal and population scope conditions, however, the properties of text-based estimation of primary candidates' positions allow for research into previously difficult-to-study phenomena. Because this measurement model makes no assumption about individuals' continuity across elections, candidates' positions can be tracked over time and space (in cases of, for instance, redistricting or opting to run in a new district). The word-level parameters included in scaling results, representing each term's general sample frequency and ability to discriminate between positions, also highlight the potential to systematically investigate the substance of House primary discourse election-to-election. And while the average primary voter may not take the time to seek out their House candidates' campaign websites, the plain-English campaign platforms present an opportunity to evaluate whether citizens' perceptions of candidates' positions track the scaling algorithm-produced estimates or not.

Perhaps most promisingly, estimates of candidates' positions measured independently

of their campaign contributions and (incumbents') roll-call votes suggests promising new avenues of study. As highlighted by primary campaign positions' far-from-perfect correlations with DW-NOMINATE and CF Scores, each measures something different. The ability to estimate campaign positioning without having to assume that receipts or legislative voting *are* candidates' positions opens the door to meaningful investigation into the substantive relationships between public campaign rhetoric, support from moneyed interests, and subsequent legislative behavior.

References

Abramowitz, Alan I., and Steven Webster. 2016. "The rise of negative partisanship and the nationalization of U.S. elections in the 21st century." *Electoral Studies* 41: 12–22.

Abramowitz, Alan I., Brad Alexander, and Matthew Gunning. 2006. "Incumbency, Redistricting, and the Decline of Competition in U.S. House Elections." *Journal of Politics* 68(1): 75–88.

Ansolabehere, Stephen, James M. Snyder, and Charles Stewart. 2001. "Candidate Positioning in U.S. House Elections." *American Journal of Political Science* 45(1): 136–159.

Bafumi, Joseph, and Michael C. Herron. 2010. "Leapfrog Representation and Extremism: A Study of American Voters and Their Members in Congress." *American Political Science Review* 104(3): 519–542.

Banks, Jeffrey S. 1990. "A model of electoral competition with incomplete information." *Journal of Economic Theory* 50(2): 309–325.

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

Barber, Michael. 2022. "Comparing Campaign Finance and Vote-Based Measures of Ideology." *The Journal of Politics* 84(1): 613–619.

Barber, Michael J., Brandice Canes-Wrone, and Sharece Thrower. 2017. "Ideologically Sophisticated Donors: Which Candidates Do Individual Contributors Finance?" *American Journal of Political Science* 61(2): 271–288.

Baron, David P. 1994. "Electoral Competition with Informed and Uniformed Voters." *American Political Science Review* 88(1): 33–47.

Bartels, Larry M. 1986. "Issue Voting Under Uncertainty: An Empirical Test." *American Journal of Political Science* 30(4): 709–728.

Bonica, Adam. 2013. "Ideology and Interests in the Political Marketplace." *American Journal of Political Science* 57(2): 294–311.

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

Bonica, Adam. 2019. "Database on Ideology, Money in Politics, and Elections (DIME): Public version 3.0 [Computer file]. https://web.stanford.edu/~bonica/data.html." Stanford University Libraries.

Bonica, Adam, and Gary W Cox. 2018. "Ideological Extremists in the U.S. Congress: Out of Step but Still in Office." *Quarterly Journal of Political Science* 13(1): 207–236.

Brady, David W., Hahrie Han, and Jeremy C. Pope. 2007. "Primary Elections and Candidate Ideology: Out of Step with the Primary Electorate?" *Legislative Studies Quarterly* 32(1): 79–105.

Cameron, Charles M, and James M Enelow. 1992. "Asymmetric policy effects, campaign contributions, and the spatial theory of elections." *Mathematical and Computer Modelling* 16(8-9): 117–132.

Canes-Wrone, Brandice, David W. Brady, and John F. Cogan. 2002. "Out of Step, Out of Office: Electoral Accountability and House Members' Voting." *American Political Science Review* 96(1): 127–140.

Clinton, Joshua D. 2012. "Using Roll Call Estimates to Test Models of Politics." *Annual Review of Political Science* 15(1): 79–99.

Clinton, Joshua D., and Adam Meirowitz. 2001. "Agenda Constrained Legislator Ideal Points and the Spatial Voting Model." *Political Analysis* 9(3): 242–259.

Clinton, Joshua D., and John Lapinski. 2008. "Laws and Roll Calls in the U.S. Congress, 1891–1994." *Legislative Studies Quarterly* 33(4): 511–541.

Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98(2): 355–370.

Cowburn, Mike, and Marius Sältzer. N.d. "The Effect of Primaries on Intra-Campaign Positional Shifts in Congressional." : Forthcoming.

Cox, Gary W., and Mathew D. McCubbins. 2005. *Setting the Agenda: Responsible Party Government in the U.S. House of Representatives*. Cambridge University Press.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper.

Druckman, James N., Cari Lynn Hennessy, Martin J. Kifer, and Michael Parkin. 2010. "Issue Engagement on Congressional Candidate Web Sites, 2002—2006." *Social Science Computer Review* 28(1): 3–23.

Druckman, James N., Martin J. Kifer, and Michael Parkin. 2007. "The Technological Development of Congressional Candidate Web Sites: How and Why Candidates Use Web Innovations." *Social Science Computer Review* 25(4): 425–442.

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.

Egerod, Benjamin C.K., and Robert Klemmensen. 2020. "Scaling Political Positions from Text: Assumptions, Methods and Pitfalls." In *The SAGE Handbook of Research Methods in Political Science and International Relations*. SAGE Publications , 498–521.

Enelow, James M., and Melvin J. Hinich. 1982. "Ideology, Issues, and the Spatial Theory of Elections." *American Political Science Review* 76(3): 493–501.

Enelow, James M., and Melvin J. Hinich. 1984. *The Spatial Theory of Voting: An Introduction*. Cambridge University Press.

Gimpel, James G., Frances E. Lee, and Shanna Pearson-Merkowitz. 2008. "The Check Is in the Mail: Interdistrict Funding Flows in Congressional Elections." *American Journal of Political Science* 52(2): 373–394.

Grimmer, Justin. 2013. "Appropriators not Position Takers: The Distorting Effects of Electoral Incentives on Congressional Representation." *American Journal of Political Science* 57(3): 624–642.

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.

Hall, Andrew B. 2015. "What Happens When Extremists Win Primaries?" *American Political Science Review* 109(1): 18–42.

Hall, Andrew B, and James M Snyder. 2015. "Candidate Ideology and Electoral Success." : Forthcoming.

Hassell, Hans J.G. 2023. "Party Elite Engagement and Coordination in House Primary Elections: A Test of Theories of Parties." *American Journal of Political Science* 67(2): 307–323.

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." *The Journal of Politics* 84(3): 1760–1776.

Herrnson, Paul S., Costas Panagopoulos, and Kendall L. Bailey. 2020. *Congressional Elections: Campaigning at Home and in Washington*. CQ Press.

Hirano, Shigeo, and James M. Snyder. 2019. *Primary Elections in the United States*. Cambridge: Cambridge University Press.

Hopkins, Daniel J. 2018. *The Increasingly United States: How and Why American Political Behavior Nationalized*. 1 edition ed. Chicago: University of Chicago Press.

Jacobson, Gary C. 1990. "The Effects of Campaign Spending in House Elections: New Evidence for Old Arguments." *American Journal of Political Science* 34(2): 334–362.

Jacobson, Gary C. 2015. "It's Nothing Personal: The Decline of the Incumbency Advantage in US House Elections." *Journal of Politics* 77(3): 861–873.

Kujala, Jordan. 2020. "Donors, Primary Elections, and Polarization in the United States." *American Journal of Political Science* 64(3): 587–602.

Lachat, Romain. 2011. "Electoral Competitiveness and Issue Voting." *Political Behavior* 33(4): 645–663.

Lauderdale, Benjamin E., and Alexander Herzog. 2016. "Measuring Political Positions from Legislative Speech." *Political Analysis* 24(3): 374–394.

Laver, Michael, Kenneth Benoit, and John Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 97(02).

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

Lewis, Jeffrey B., Keith Poole, Howard Rosenthal, Aaron Rudkin, Adam Boche, and Luke Sonnet. 2023. "Voteview: Congressional Roll-Call Votes Database.".

Lewis, Jeffrey B, and Luke Sonnet. N.d. "Estimating NOMINATE scores over time using penalized splines." : Forthcoming.

Lockhart, Mackenzie, and Seth J. Hill. N.d. "How Do General Election Incentives Affect the Visible and Invisible Primary?" *Legislative Studies Quarterly*: Forthcoming.

Lowe, Will, Kenneth Benoit, Slava Mikhaylov, and Michael Laver. 2011. "Scaling Policy Preferences from Coded Political Texts." *Legislative Studies Quarterly* 36(1): 123–155.

Martin, Gregory J., and Joshua McCrain. 2019. "Local News and National Politics." *American Political Science Review* 113(2): 372–384.

McCarty, Nolan, Keith Poole, and Howard Rosenthal. 2006. *Polarized America: The Dance of Ideology and Unequal Riches*. MIT Press.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. 2016. *Polarized America: The Dance of Ideology and Unequal Riches*. MIT Press.

McDonald, Maura, Rachel Porter, and Sarah A. Treul. 2020. "Running as a Woman? Candidate Presentation in the 2018 Midterms." *Political Research Quarterly* 73(4): 967–987.

McGhee, Eric, Seth Masket, Boris Shor, Steven Rogers, and Nolan McCarty. 2014. "A Primary Cause of Partisanship? Nomination Systems and Legislator Ideology." *American Journal of Political Science* 58(2): 337–351.

Meirowitz, Adam. 2005. "Informational Party Primaries and Strategic Ambiguity." *Journal of Theoretical Politics* 17(1): 107–136.

Meisels, Mellissa, Joshua D Clinton, and Gregory A. Huber. N.d. "Giving to the Extreme? Experimental Evidence on Donor Response to Candidate and District Characteristics." : Forthcoming.

Milita, Kerri, John Barry Ryan, and Elizabeth N. Simas. 2014. "Nothing to Hide, Nowhere to Run, or Nothing to Lose: Candidate Position-Taking in Congressional Elections." *Political Behavior* 36(2): 427–449.

Miller, Warren E., and Donald E. Stokes. 1963. "Constituency Influence in Congress." *American Political Science Review* 57(1): 45–56.

Moskowitz, Daniel J. 2021. "Local News, Information, and the Nationalization of U.S. Elections." *American Political Science Review* 115(1): 114–129.

Nokken, Timothy P., and Keith T. Poole. 2004. "Congressional Party Defection in American History." *Legislative Studies Quarterly* 29(4): 545–568.

Nyhan, Brendan, and Jacob M. Montgomery. 2015. "Connecting the Candidates: Consultant Networks and the Diffusion of Campaign Strategy in American Congressional Elections." *American Journal of Political Science* 59(2): 292–308.

Patty, John W., and Elizabeth Maggie Penn. 2019. "Are Moderates Better Representatives than Extremists? A Theory of Indirect Representation." *American Political Science Review* 113(3): 743–761.

Phillips, Connor Halloran, James M Snyder, and Andrew B Hall. N.d. "Who Runs for Congress? A Study of State Legislators and Congressional Polarization." : Forthcoming.

Poole, Keith, Jeffrey B. Lewis, James Lo, and Royce Carroll. 2011. "Scaling Roll Call Votes with wnominate in R." *Journal of Statistical Software* 42(1): 1–21.

Poole, Keith T. 2005. "Spatial Models of Parliamentary Voting.".

Poole, Keith T., and Howard Rosenthal. 1991. "Patterns of Congressional Voting." *American Journal of Political Science* 35(1): 228–278.

Porter, Rachel, Sarah A Treul, and Maura McDonald. N.d. "Changing the Dialogue: Descriptive Candidacies & Position Taking in Campaigns for the U.S. House of Representatives." *Journal of Politics*: Forthcoming.

Rheault, Ludovic, and Christopher Cochrane. 2020. "Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora." *Political Analysis* 28(1): 112–133.

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.

Stuckatz, Jan. 2022. "How the Workplace Affects Employee Political Contributions." *American Political Science Review* 116(1): 54–69.

Sulkin, Tracy, Cortney M. Moriarty, and Veronica Hefner. 2007. "Congressional Candidates' Issue Agendas On- and Off-line." *Harvard 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." *The Journal of Politics* 75(2): 330–342.

Temporão, Mickael, Corentin Vande Kerckhove, Clifton Van Der Linden, Yannick Dufresne, and Julien M. Hendrickx. 2018. "Ideological Scaling of Social Media Users: A Dynamic Lexicon Approach." *Political Analysis* 26(4): 457–473.

Theriault, Sean M. 2006. "Party Polarization in the US Congress: Member Replacement and Member Adaptation." *Party Politics* 12(4): 483–503.

Thomsen, Danielle M. 2014. "Ideological Moderates Won't Run: How Party Fit Matters for Partisan Polarization in Congress." *Journal of Politics* 76(3): 786–797.

Thomsen, Danielle M. 2022. "Competition in Congressional Elections: Money versus Votes." *American Political Science Review*: Forthcoming.

Utych, Stephen M. 2020. "Man Bites Blue Dog: Are Moderates Really More Electable than Ideologues?" *Journal of Politics* 82(1): 392–396.

Vafa, Keyon, Suresh Naidu, and David Blei. 2020. "Text-Based Ideal Points." Paper presented at the Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online.

Warshaw, Christopher, and Chris Tausanovitch. 2022. "Subnational Ideology and Presidential Vote Estimates (v2022)." Harvard Dataverse.

Woon, Jonathan. 2018. "Primaries and Candidate Polarization: Behavioral Theory and Experimental Evidence." *American Political Science Review* 112(4): 826–843.

Appendix

Contents

A Data Collection Details

Identifying relevant candidates. I used Ballotpedia.com to identify all candidates who appeared on a Republican or Democratic primary ballot in each district in 2016, 2018, 2020, and 2022, as well as take down the primary election date and candidate type (incumbent/open seat/challenger). Independent, write-in, and dropout candidates were excluded, as well as candidates who ran in the primaries in the table below.

Table A1. Excluded Primary Races

Locale	Reason
Alaska, 2022 only	Top-4
California	Top-2
Connecticut	Party Convention
Louisiana	Top-2
Utah	Party Convention
Virginia, 2016, Democratic: Districts 5,7,1,6,9,10	Party Convention
Virginia, 2016, Republican: Districts 3,8,5,11,7	Party Convention
Virginia, 2018, Democratic: District 5	Party Convention
Virginia, 2018, Republican: District 5,8,3,7,6	Party Convention
Virginia, 2020, Democratic: District 9	Party Convention
Virginia, 2020, Republican: District 8,5,10,11,4,7	Party Convention
Virginia, 2022, Republican: District 8,5,10,11	Party Convention
Washington	Top-2

Source: Footnotes of FEC primary date calendars.

Searching for campaign websites in real time. Data on 2022 primary candidates were collected in real time. Candidates' web pages were accessed as immediately as possible before their primary, always within a week of the election date. I first performed a web search for "[candidate name] for Congress [election year]". Official governmental websites and social media sites were ignored. If no website appearing to be the candidate's campaign website appeared in the first page of search results, I added the district (e.g. "AL-1") to the search terms. If nothing appeared, I then consulted Politics1.com and Ballotpedia.com, which compile fairly reliable lists of candidates' campaign websites at various levels of government. If no non-social media website or non-governmental campaign website was found, I moved on to the next candidate. Although it is possible that some candidate websites eluded this data collection process, websites that were not found while deliberating searching via numerous steps were not readily accessible to members of the public, activists, or journalists, who would almost certainly devote less effort to find them.

Searching for archived campaign websites. For candidates who ran in 2016, 2018, and 2020, the process was identical to that outlined above, with an added step of accessing the archived website as it appeared at the relevant time via the Wayback Machine (archive.org). I first performed a web site for "[candidate name] for Congress [election year]". Some candidates ran in more recent elections and maintained a new website at the same URL which hosted their campaign website during the election year of interest. Because many candidates delete their campaign websites after losing election, I likewise consulted historic versions of Politics1.com and Ballotpedia.com. Once a potential historic campaign website URL was identified, I pasted it into the Wayback Machine and accessed the snapshot of the website most immediately before the date of the primary. While these archives ranged in time from very close to the primary to months before the primary, I also recorded the date of the archive version.

Identifying issue positions. The vast majority of campaign websites had clearly delineated pages or sections for policy platforms, issue positions, or candidate priorities. If the area devoted to positions was not readily obvious in the website architecture, I surveyed the entirety of the website for other places where one might find issue positions. I do not consider candidate biographies, endorsement lists, campaign updates, or volunteer/donation pages to be issue positions. Many incumbent candidates (and some candidates with state legislative experience) devoted a section of the website to their legislative achievements, and these were nearly always separate from issue position pages. I excluded pages devoted exclusively to legislative achievements, but some candidates relate positions on their issue pages to legislative achievements, all of which I include as issue positions. If a campaign website with issue position content was successfully accessed, the URL was recorded in a spreadsheet.

Collecting issue position text. Once issue position content was identified, I manually copied and pasted all of the associated positioning text — including the section header, issue stances, and candidate quotes — from each sub-issue page or section into one .txt file titled the candidate's name and election year. I also captured the website content exactly as it appeared with a combination of manual screen capture and automated screen capture via the Awesome Screenshot extension on Google Chrome.

B   Technical Scaling Details

B.1   Text Processing Flow

To prepare the text of primary candidates' issue positions for scaling, I build a corpus of documents, or a collection of all individual primary campaign platforms. I then tokenize each document's text with terms standardized to all-lowercase and remove punctuation. Next, I preserve key non-unigram phrases found by compounding the separate terms.

To improve computing performance, I remove "stop words" such as "and", "for", and "of", which are used very frequently and provide negligible substantive information. I then reduce terms to their stems in order to combine terms that have the same central meaning yet slightly different suffixes and prefixes — for example, "reduce", "reduction", and "reducing" share the stem "reduc".

When utilizing unsupervised scaling methods, it is important to ensure that the dimension of interest — here, a left-right, issue-based dimension — is the dominant dimension structuring rhetorical discourse within the corupus. As such, it is beneficial to discard terms that are irrelevant to the dimension of interest and relevant to an orthogonal dimension within which the algorithm may get "stuck" (Grimmer and Stewart 2013; Egerod and Klemmensen 2020). I discard terms related to congressional procedure, which are overwhelmingly used by sitting legislators, as well as commonly-used geographical terms, in order to protect against identifying an incumbency-based dimension or region-based dimension. In practice, this refinement is inconsequential to the vast majority of primary candidates' position estimates as illustrated by the strong correlation between estimates with and without these terms and the non-unigrams shown in the left panel of Figure X. To improve computing time and drop other terms uninformative of the global dimension, I discard terms used in 100 campaign platforms or fewer — a lenient requirement given that the corpus consists of almost 4,000 campaign platforms.

The resulting $\mathbf{N} \times \mathbf{M}$ document-feature matrix consists of $j = 1, ..., m$ term columns, $it = 1, ..., n$ candidate-year rows, and term frequencies as cell entries.

Table B1. Scaling Refinements

Procedural Terms Dropped	Non-Unigram Terms Included
`"hr"`, `"h.r"`, `"co-chair"`,`"congresswoman"`, `"congressman"`, `"co-sponsor"`,`"reauthor"`, `"codifi"`, `"chair"`, `"caucus"`,`"introduc"`, `"passag"`, `"subcommitte"`,`"cosponsor"`, `"committe"`, `"lawmak"`, `"mayor"`, `"congress"`, `"chairman"`, `"speaker"`, `"legislatur"`, `"re-elect"`, `"hyperlink"`	`"first amendment"`, `"1st amendment"`, `"second amendment"`, `"2nd amendment"`, `"planned parenthood"`, `"right to bear arms"`, `"mandatory minimum"`, `"mandatory minimums"`, `"mental health"`, `"clean energy"`, `"sexual assault"`, `"student loan"`, `"student loans"`, `"sexual violence"`, `"critical race theory"`, `"religious freedom"`, `"reproductive freedom"`, `"freedom of speech"`, `"freedom of expression"`, `"freedom of religion"`, `"cancel culture"`, `"debt ceiling"`, `"balanced budget"`, `"common core"`, `"build the wall"`, `"sanctuary city"`, `"sanctuary cities"`

Note: Scaling excludes procedural terms as well as geographic terms, and includes compounded non-unigram terms.

B.2 Estimation with `wordfish`

`wordfish` (Slapin and Proksch 2008) is an unsupervised machine learning algorithm for scaling political text to infer the source's latent position on a single dimension. Based on a Poisson IRT model, `wordfish` uses an iterative expectation maximization algorithm due to the need to estimate both term-level and candidate-level parameters as a function of observed term usage.

The rate y at which primary candidate i uses term j in election year t is assumed to be drawn from a Poisson distribution, which is characterized by a single parameter λ representing both the expectation and variance. This parameter logarithmically links the probability distribution generating the observed term rate to the linear predictors of interest to be estimated:

$$y_{ijt} \sim Poisson(\lambda_{ijt})$$
$$\lambda_{ijt} = exp(\alpha_{it} + \psi_j + \beta_j * \omega_{it})$$

The key parameter is ω, which stands in for candidate i's latent primary campaign position in election t. β represents word j's weight or, put differently, its importance in discriminating between campaign positions. A word fixed effect ψ captures the rate at which word j is used in general, and a candidate-year fixed effect α captures the verbosity of candidate i's campaign position text in election t.

Parameter estimation is initialized with start values consisting of "best guesses" based upon term frequencies. Term fixed effects ψ_j begin as term j's logged average count, while the fixed effect for the first candidate-year (α_1) is set to 0 and $\alpha_{2,...,n}$ begin as the logged

average word count relative to that of $it = 1$. Start values for term weights β and candidate-year positions ω are the left and right singular vectors obtained from an SVD of the matrix of term and candidate-year residuals. Unsurprisingly, final estimates of ω correlate highly with nonparametric estimates resulting from a simpler correspondence analysis as shown in the Alternative Scalings subsection. As such, the methodology from which my primary campaign positions derive bears strong resemblance to the augmented CA methodology used for Bonica's (2014) estimates of candidate ideology.

Estimation proceeds iteratively, with term parameters ψ and β first fixed at their start values and candidate-year parameters ω and α calculated conditionally on the expected term parameters. The following conditional log-likelihood is maximized for each candidate-year:

$$\sum_{j=1}^{m}(-\lambda_{ijt} + ln(\lambda_{ijt}) * y_{ijt})$$

where

$$\lambda_{ijt} = exp(\alpha_{it} + \psi_j^{prev} + \beta_j^{prev} * \omega_{it}).$$

To identify the global directionality of candidate positions ω, a pair of documents (candidate-years) are specified with an inequality constraint. Moreover, the mean of candidate positions across all years is equal to 0 and the standard deviation is set to 1.

Taking the expected values of candidate-year parameters ω and α obtained previously, term parameters ψ and β are then calculated conditionally with the following log-likelihood maximized for each term:

$$\sum_{it=1}^{n}(-\lambda_{ijt} + ln(\lambda_{ijt}) * y_{ijt})$$

where

$$\lambda_{ijt} = exp(\alpha_{it}^{prev} + \psi_j + \beta_j * \omega_{it}^{prev}).$$

The overall log-likelihood of the model with the new parameter estimates is then calculated as the sum of the term log-likelihoods conditional upon the candidate-year log-likelihoods:

$$\sum_{j}^{m}\sum_{it=1}^{n}(-\lambda_{ijt} + ln(\lambda_{ijt}) * y_{ijt}).$$

The candidate-year parameters are then re-calculated based upon the new term parameters, and the resulting candidate-year parameters are used to repeat the term parameter calculation. The conditional maximum likelihoods are calculated iteratively until the log-posterior reaches a convergence threshold of a one-millionth and the differences in parameter values from the previous iteration are under a hundred-millionth.

B.3 Alternative Scalings

The following figures plot relationships between the main scaling specification and alternative scalings consisting of: leaving the tokens unrefined by keeping procedural and geographic terms and not non-unigrams, simple unidimensional correspondence analysis, incumbent-only scaling, and year-specific scaling. All correlations are above 0.90.

Figure B1. Relationship Between Primary Campaign Scores and Alternative Scalings
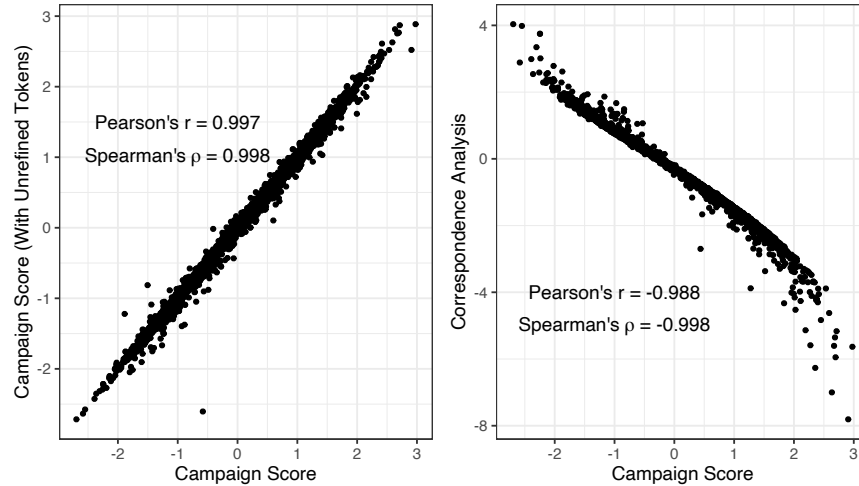
Note: Text-based scaling estimates of primary campaign positions along y-axes, estimates from unigram-only scaling including geographic and procedural terms (left) and from unidimensional correspondence analysis (right) along x-axes. Pearson and Spearman's ranking correlations show strong relationships.

Figure B2. Campaign Scores From Pooled and Incumbent-Only Scaling
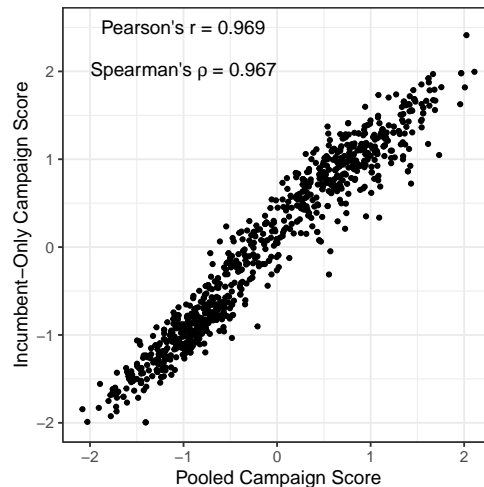
Note: Relationship between incumbents' campaign scores from pooled scaling and incumbent-only scaling. Pearson and Spearman's ranking correlations show strong relationships.

Figure B3. Relationship Between Campaign Scores From Pooled and Year-Specific Scaling
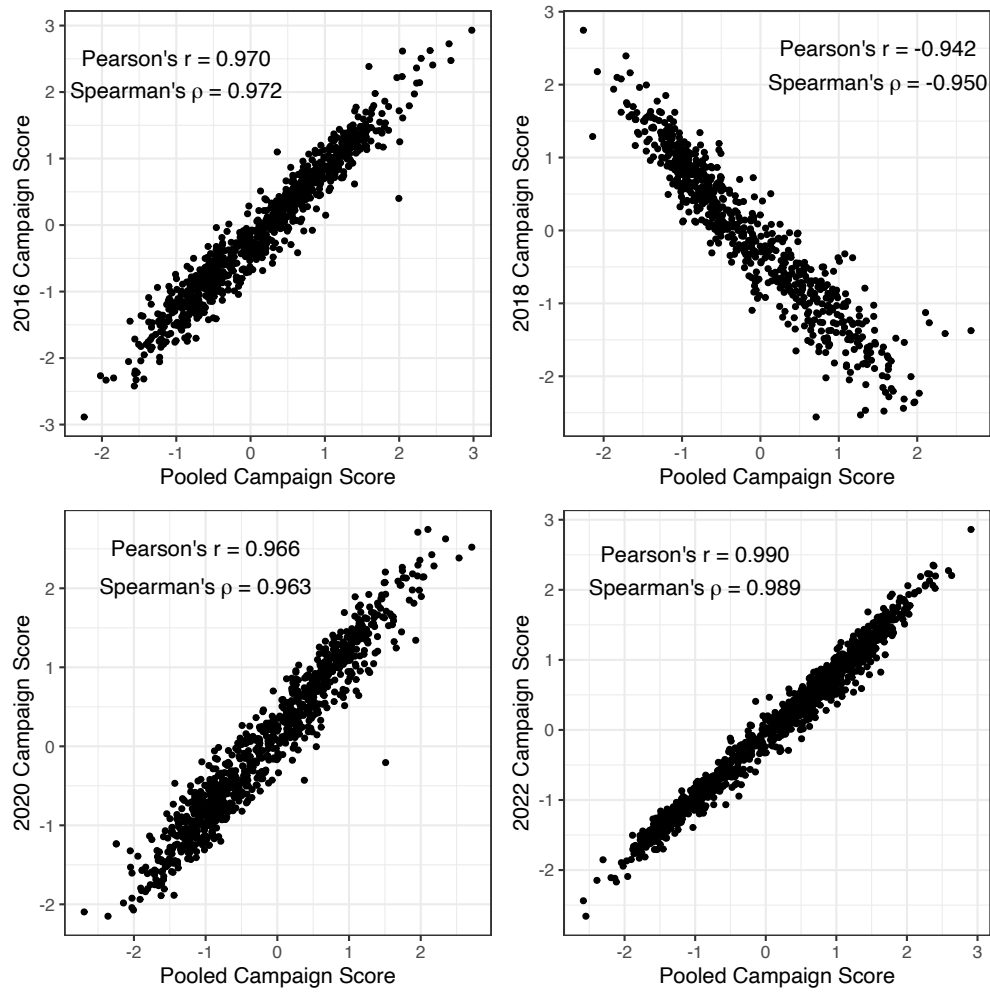
Note: Relationship between campaign scores from pooled scaling and each year scaled separately. Spearman's ranking correlations show strong relationships.

B.4 Top Discriminating Scaling Terms

Scaling results include word-level parameters β, a discrimination parameter, and ψ, an overall frequency parameter. Terms with the highest β are those that exert the greatest change to a candidate's campaign position, such that the most positive (negative) terms are most strongly associated with conservative (liberal) positions. The following tables report terms from the main pooled and year-specific scalings sorted by largest negative and positive β.

Table B2. Top 20 Most Conservative and Liberal Terms

	Conservative	β (weight)	ψ (FE)	**Liberal**	β (weight)	ψ (FE)
1	critical race theori	1.579	-3.306	community-bas	-1.990	-4.569
2	build the wal	1.566	-3.963	rental	-1.970	-4.425
3	tyrann	1.547	-3.885	equit	-1.959	-3.476
4	crt	1.527	-3.630	reproduct	-1.733	-2.395
5	indoctrin	1.510	-3.320	trauma	-1.713	-3.993
6	god	1.469	-1.869	matern	-1.678	-3.375
7	tyranni	1.466	-3.307	lgbtq	-1.637	-2.214
8	christian	1.453	-2.680	high-capac	-1.630	-4.550
9	sanctiti	1.447	-2.997	lewi	-1.621	-4.076
10	god-given	1.443	-3.345	low-incom	-1.586	-2.608
11	unborn	1.441	-1.930	dispar	-1.578	-3.093
12	pro-lif	1.436	-1.684	childcar	-1.572	-2.865
13	communist	1.388	-2.924	disproportion	-1.558	-2.780
14	swamp	1.373	-3.338	pell	-1.555	-3.514
15	socialist	1.365	-2.721	pre-k	-1.541	-3.105
16	amnesti	1.344	-2.393	tuition-fre	-1.506	-4.158
17	islam	1.336	-2.572	underserv	-1.467	-3.464
18	sanctuary c	1.333	-2.773	expung	-1.461	-4.104
19	alien	1.320	-2.301	resili	-1.386	-3.318
20	2nd amend	1.315	-1.483	discriminatori	-1.383	-3.526

Note: Top terms with most positive (conservative) discrimination parameters and most negative (liberal) discrimination parameters.

Table B3. Top 20 Most Liberal Terms by Year

	2016	2018	2020	2022
1	student_loan	student_loan	lgbtq	equit
2	colleg	reproduct	racial	reproduct
3	infrastructur	mental_health	reproduct	low-incom
4	senior	epidem	disproportion	lewi
5	workforc	color	inequ	childcar
6	mental_health	violenc	incarcer	pre-k
7	bridg	childhood	+	good-pay
8	earli	priorit	low-incom	bargain
9	discrimin	clean_energi	color	inequ
10	loan	communiti	orient	lgbtq
11	invest	opioid	black	disproportion
12	partnership	transport	gender	high-qual
13	climat	champion	workplac	clean_energi
14	minimum	rural	discrimin	incarcer
15	univers	student	justic	childhood
16	access	access	emiss	gap
17	transit	transit	sexual	workplac
18	violenc	expand	fossil	discrimin
19	graduat	prescript	prison	climat
20	student	21st	gap	black

Table B4. Top 20 Most Conservative Terms by Year

	2016	2018	2020	2022
1	pro-lif	properti	pro-lif	critical_race_theori
2	amnesti	said	unborn	crt
3	unborn	liberti	balanced_budget	indoctrin
4	2nd_amend	obamacar	2nd_amend	god
5	liberti	2nd_amend	shall	unborn
6	constitut	bureaucrat	infring	pro-lif
7	common_cor	bear	concept	pelosi
8	second_amend	constitut	bureaucrat	communist
9	balanced_budget	illeg	obamacar	finish
10	illeg	answer	second_amend	overreach
11	bear	principl	illeg	liber
12	ir	say	radic	speech
13	concept	religi	border	infring
14	obamacar	second_amend	liberti	radic
15	border	man	bear	2nd_amend
16	epa	spend	southern	right_to_bear_arm
17	faith	govern	china	second_amend
18	bureaucrat	abort	presid	concept
19	principl	term	constitut	shall
20	repeal	deficit	conserv	ideolog

C   Application Robustness

The following figures and tables demonstrate the robustness of the results presented in the brief application of the new measure. The figures show that the trends from Figure 4 hold when subsetting to 2016 and 2018 candidates with CF Scores and when fitting a Loess curve instead of a straight line. The tables show that the main results do not mask considerable heterogeneity in the effect of district partisanship by candidate type and using Tausanovitch and Warshaw's (2013) updated district ideology (conservatism) MRP estimates from (Warshaw and Tausanovitch 2022) leads to similar results. However, this measure's mapping onto the two-year House election time periods is even more problematic than presidential vote share: it is broken down into surveys from 2012-2016 and 2017-2021, so the former is matched to 2016 candidates, while the latter is matched to 2018, 2020, and 2022 candidates even though the surveys used to construct the measure do not include 2022. Although this variable is scaled to have a universe-wide mean 0 SD 1, it only ranges from –0.5 to 0.4 in House districts during this time period, so I rescale the variable such that a 0.1 increase constitutes a one-unit increase in the regression.
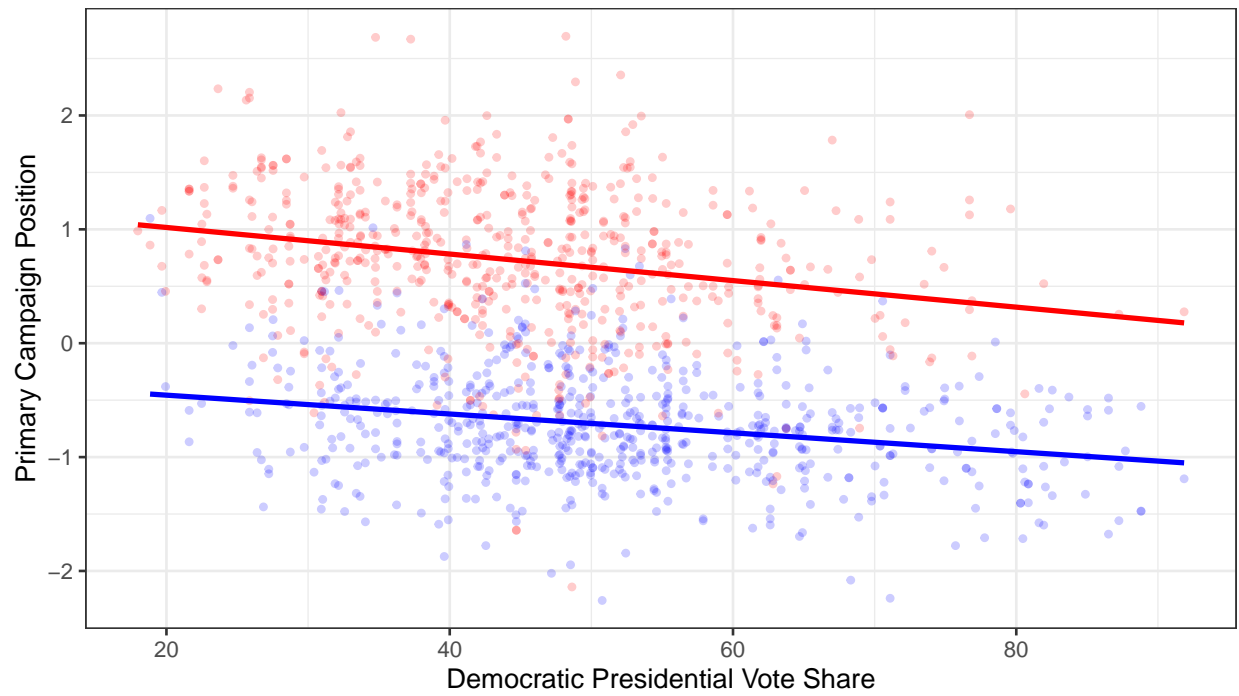
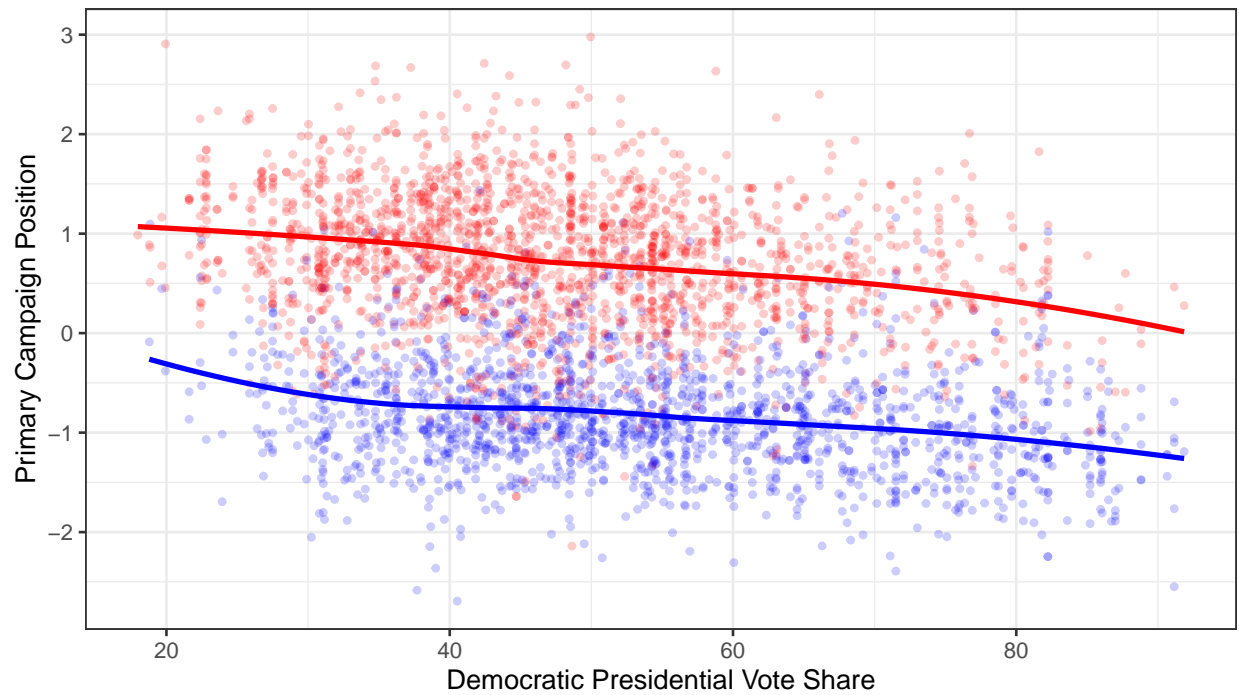**Figure C1.** Only Candidates With CF Scores

Figure C2. With Loess Curves Fit

Table C1. District Partisanship and Candidate Positions With Interactions

	Primary Campaign Position		Recipient CF Score	
	Democrats	Republicans	Democrats	Republicans
(Intercept)	−0.562***	0.465***	−0.702***	0.985***
	(0.043)	(0.049)	(0.049)	(0.024)
District Dem. Partisanship	−0.010***	−0.017***	0.001	−0.003***
	(0.002)	(0.003)	(0.002)	(0.001)
Open Seat Candidate	0.0005	0.358***	−0.535***	0.151*
	(0.046)	(0.055)	(0.062)	(0.065)
Primary Challenger	0.214*	0.636***	−0.555***	0.332**
	(0.085)	(0.071)	(0.167)	(0.107)
General Challenger	0.045	0.347***	−0.649***	0.280***
	(0.045)	(0.059)	(0.053)	(0.061)
District * Open	−0.001	−0.002	0.007	−0.004
	(0.003)	(0.004)	(0.003)	(0.005)
District * Prim. Chall.	−0.0005	0.016**	0.008	0.007
	(0.004)	(0.005)	(0.007)	(0.008)
District * Gen. Chall.	0.005	0.004	−0.003	−0.0001
	(0.003)	(0.004)	(0.003)	(0.006)
Year Fixed Effects	✓	✓	✓	✓
Observations	1,740	1,995	1,090	953
Adjusted R^2	0.138	0.161	0.209	0.023

Note: $^*p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

<div align="center">

Table C2. District Ideology and Candidate Positions

</div>

	Primary Campaign Position		Recipient CF Score	
	Democrats	Republicans	Democrats	Republicans
(Intercept)	−0.630***	0.373***	−0.693***	0.918***
	(0.033)	(0.041)	(0.033)	(0.040)
District Conservatism	0.071***	0.140***	−0.009	0.058**
	(0.010)	(0.011)	(0.011)	(0.020)
Open Seat Candidate	0.012	0.393***	−0.519***	0.195***
	(0.035)	(0.040)	(0.048)	(0.046)
Primary Challenger	0.188***	0.454***	−0.386***	0.258**
	(0.044)	(0.042)	(0.076)	(0.081)
General Challenger	0.041	0.369***	−0.632***	0.338***
	(0.038)	(0.047)	(0.041)	(0.065)
Year Fixed Effects	✓	✓	✓	✓
Observations	1,737	1,986	1,090	953
Adjusted R^2	0.131	0.167	0.201	0.031

Note: *p<0.05; **p<0.01; ***p<0.001