Measuring Strategic Positioning in Congressional Elections

Colin R. Case*

August 4, 2023

Abstract

How do incumbents respond to extreme primary challengers? While theoretical models suggest incumbents should adopt more extreme positions, previous research lacks support for this hypothesis. However, data limitations have forced researchers to focus only on legislative behavior and not consider positioning during the campaign in response to primary challengers. To overcome these data limitations, I introduce Website Em-Bedding (WEB) Strategic Positioning Scores. WEB Scores employ word embeddings with document-level vectors trained on congressional candidates' issue statements, as presented on their campaign websites. These estimates have high construct validity and improve upon current measurement limitations, including expanding the number of candidates with estimates and using actual issue-position data to produce these estimates. Consistent with theoretical expectations, I show incumbent candidates become more extreme (moderate) in their issue positioning during the campaign in response to an extreme (moderate) primary challenger whereas previous measures do not.

*The University of North Carolina at Chapel Hill, Department of Political Science. I am grateful for helpful comments and feedback from Max Goplerud, Eric Hansen, Jacob Montgomery, Santiago Olivella, Emily Ommundsen, Rachel Porter, Jason Roberts, Tim Ryan, Brandon Stewart, and Sarah Treul. I would also like to thank Adam Bonica for his willingness to share CFscores for the 2020 election. Previous versions of this paper were presented at the 2022 Annual Meeting for the Society of Political Methodology and the 2023 Junior Americanist Workshop Series MPSA Mini-Conference – I am grateful for participants' engagement and feedback.

In 2020, Representative Andy Kim ran unopposed in the Democratic primary for New Jersey's 3rd congressional district. Rep. Kim advocated for moderate policies for the Democratic Party, suggesting "pragmatic" solutions to address climate change, "bipartisan" proposals to rebuild America's infrastructure, and "building" on the Affordable Care Act. In 2022, however, Kim was challenged in the Democratic primary by Reuven Hendler, a first-time candidate running for office. Hendler ran far to the left of Kim, strongly advocating for a Medicare-for-all healthcare system, tuition-free college, and aggressive policies to combat climate change. In response to this extreme challenger, Kim shifted his own positions to the left; he proceeded to win the primary decisively, with almost 93% of the vote. Despite theoretical expectations supporting Kim's behavior, prior research lacks evidence that an extreme primary challenger is associated with a change in aggregate incumbent positioning (e.g., Pearson and Lawless 2008; Boatright 2014; Hirano and Snyder 2019).

Why has prior research found a lack of support for incumbents responding to extreme challengers? I argue measurement limitations have prevented scholars from fully testing how incumbents position themselves in response to primary challengers' positioning. The aforementioned research places a substantive focus on roll-call behavior and does not consider alternative changes, such as the issues candidates run on, in response to an extreme primary challenger. The ability to test an incumbent's response, outside of legislative behavior, to a primary challenger's positioning has been limited by existing measures for two reasons. First, measures of candidate positioning often exclude a large proportion of challengers. For example, because Hendler did not receive any campaign contributions, did not make it to the general election stage, and did not have prior elected experience, existing measures of candidate positioning (e.g., Bonica 2014; Christopher et al. 2015; Barberá 2015; Macdonald et al. 2022; Gaynor et al. 2022) do not have an estimate for him during the election. This prevents researchers from classifying candidates like Hendler as either extreme or moderate, and therefore, unable to evaluate an incumbent's response. Second, existing measures use citizen behavior to scale candidates' positioning rather than measuring a candidate's positioning itself. Changes in scores using approximations could be due in part to groups of citizens', such as campaign contributors, responding to changes in electoral dynamics rather than changes in candidate behavior (Hirano and Snyder 2019). Both limitations have contributed to scholars' inability to assess how incumbents position themselves in response to primary challengers' positioning during the campaign.

Given the shortcomings described above, I propose a new measure of congressional candidate positioning, Website EmBedding (WEB) Strategic Positioning Scores, using data collected from campaign website issue pages by Porter, Treul and Case (2023) for the 2018, 2020, and 2022 primaries for the U.S. House of Representatives. Campaign websites are well-situated to study candidate positioning during the campaign: they are unmediated, in that they come directly from the campaign; not subject to other gatekeeping (e.g., media); and contain a range of policy areas (Druckman, Kifer and Parkin 2009). Moreover, campaign websites mitigate the limitations presented above by providing better coverage than existing measures. In addition, they also cover a substantively important concept of interest related to candidate positioning – the issues candidates *actually* run on.

To estimate WEB Scores, I rely on recent developments in word embedding models that allow for the inclusion of a document-level vector for each candidate-year occurrence. Rheault and Cochrane (2020) validate this approach across various contexts as a suitable way to uncover elite positioning. After estimating WEB Scores for primary candidates, I show the resulting measure has high validity when compared with related constructs and also captures important relationships in the underlying text related to candidate positioning. Using this measure, I show incumbent candidates challenged by an extreme (moderate) challenger become more extreme (moderate) in their positioning during campaigns when using WEB Scores. Consistent with prior work and my theoretical expectations, I do not find the same relationship holds when using proxy measures for candidate positioning, such as CFscores.

The contributions of this paper are as follows: first, I provide important clarification to an existing debate regarding the role of primary challengers in shaping incumbent behavior. While incumbents do not change their legislative behavior in aggregated measures such as

NOMINATE, they do change their campaign behavior; while not directly tied to legislating, this campaign behavior still has implications for future legislative action (Sulkin 2009). Second, I provide an off-the-shelf measure of candidate positioning for researchers that increases the number of candidates covered as well as actually captures the issues that candidates run on. This measure presents expanded possibilities for applied researchers interested in the role of candidate positioning in congressional elections. In the same vein, I also offer guidance for researchers to consider when choosing a measure of candidate positioning. Finally, the rich nature of the resulting candidate and word embeddings presents new research opportunities for studying rhetoric in congressional elections that I demonstrate in the paper.

Partisan Polarization and Primary Elections

Prior research has assessed the extent to which primary elections contribute to elite polarization through a number of different mechanisms, from extreme candidates' likelihood of success in primary elections (e.g., Hall and Snyder 2015; Thomsen 2020), to primary voters' preferences for extreme candidates (e.g., Brady, Han and Pope 2007; Jacobson 2012; Sides et al. 2020), to the higher likelihood of extreme candidates to actually run (Thomsen 2014). One avenue of particular attention has looked at whether or not incumbents respond to extreme primary challengers. There are theoretical reasons to expect incumbents should alter their behavior to maximize their electoral chances (Downs 1957). A number of scholars have tested this claim; Boatright (2014) looks at a number of different types of primary challenges and finds that none contribute to meaningful changes in positioning. Similarly, Hirano and Snyder (2019) find that there is not a meaningful difference between the proportion of incumbents moving to the extreme when looking at those who face an extreme challenger in the primary versus those who do not. It should be noted that Jewitt and Treul (2019) do find differential effects by party; those in the majority party vote less often with their party while those in the minority party vote with their party more often after facing an extreme challenger.

One of the commonalities across research assessing whether or not incumbents respond to extreme challengers is the focus on legislative behavior. In most instances, the dependent variable is a variation of NOMINATE, a scaling method that focuses on legislative position taking for select issues that make it to the floor. There is reason to suspect researchers would fail to find evidence that incumbents change their legislative positioning in response to extreme challengers. First, the presence of an extreme primary challenger suggests there are issue divisions within the party. The majority party is not incentivized to bring issues to a vote that divide the party (Cox and McCubbins 2005). Therefore, even if incumbents face an extreme challenger and change their issue positioning in response, it is unlikely incumbents would actually vote on legislation related to these issues that divide the party. Moreover, for incumbents to actually receive an electoral benefit, voters must be aware of these changes in legislative behavior. However, voters seldom know the aggregate roll-call decisions of their members within the party (Ansolabehere and Jones 2010). Furthermore, it is likely that changes in legislative behavior in response to an extreme primary challenger are different by majority party status, causing results to wash out in aggregate roll-call measures like NOMINATE (Jewitt and Treul 2019). This provides little reason to expect an effect of a primary challenger on aggregate measures of legislative position taking.

Alternatively, I argue that instead of changing their legislative positioning, incumbents change their campaign positioning and take on more extreme (moderate) positions in response to an extreme (moderate) challenger. Formal theories argue that voters prefer candidates proximal to their position (Black 1948; Downs 1957). In this manner, candidates have an incentive to take into consideration the positions of both their competitors and the electorate when deciding where to position themselves on the issues. Further, campaign positioning differs from legislative positioning in important ways. Primarily, candidates are not constrained by the legislative agenda and can talk about the issues they would like and in a nuanced way.

Prior work supports the notion that during campaigns, candidates do engage in this strategic positioning on individual issues in response to electoral factors. For example, (Sulkin 2005) find that incumbents take up new issues in response to the issue positions of their opponents. Other work highlights how the presence of candidates with certain backgrounds in primary elections, such as women or veterans, affects the issues their opponents address during the campaign (Porter, McDonald and Treul 2021). Given the strategic nature of candidates, specifically incumbents, in their issue positioning, it is likely they behave rationally and adopt more extreme (moderate) positions in the aggregate in response to extreme (moderate) challengers. Despite this, measurement limitations have prevented scholars from adequately testing this hypothesis. Generally speaking, researchers have developed a number of measurement strategies to capture candidate positioning in congressional elections. These election-focused measurement strategies that approximate candidate positioning rely on one of two sources: citizen perceptions of candidates or actual candidate behavior.

Within the first typology, measures based on citizen perceptions rely on the assumption that citizens are able to take into account candidate positioning, such as the issues they run on, policy goals, and values (Bonica 2014). Common approaches often ask survey respondents (Christopher et al. 2015; Ramey 2016) or experts (Hirano et al. 2015) to place candidates spatially from liberal to conservative and aggregate these responses to position candidates using various scaling methods. Other approaches rely on aggregate citizen behavior, such as donations (Bonica 2014) or follows on Twitter (Barberá 2015); these estimation strategies assume that citizens donate to and follow candidates on Twitter who are positioned proximal to one another.

The second typology of measurement approaches estimates candidate positioning using other related candidate behaviors. For example, Macdonald et al. (2022) use news story domain sharing (e.g., Fox News or CNN) on Twitter for members of Congress to spatially place candidates. Other approaches, such as that used by Gaynor et al. (2022), employ text-based scaling of members of Congress across a variety of different contexts, including tweets and floor speeches. Another subset of measurement strategies focuses specifically on state legislators. For example, Ansolabehere, Snyder and Stewart (2001) and Montagnes and Rogowski (2015) use Project Vote Smart's NPAT survey of state legislators while Shor and

McCarty (2011) rely on roll-call votes in state legislative bodies and use the NPAT survey to link state legislators across institutional contexts.

Across both measurement typologies, two issues persist for studying incumbents' response to primary challengers. First, existing measures exclude large populations of candidates, including many primary challengers. For example, measures that rely on survey responses from voters are often limited to general-election candidates. This is due, in part, to resource constraints: asking about over 2,000 candidates who run in congressional primaries is not feasible. There are knowledge limitations as well; it is unlikely the average voter is aware of the positions of all candidates running in a primary race (see Ahler, Citrin and Lenz 2016). Similarly, measures employing experts to place candidates from liberal to conservative often focus on general election candidates or certain high-profile races due to the same resource and knowledge constraints of experts. Other measures, such as those using on congressional floor speeches, are limited to incumbent candidates.[1]

Other measurement approaches, such as CFscores (Bonica 2014) or state legislator scores (Shor and McCarty 2011), do include a subset of primary candidates, but are still likely to exclude a large proportion of candidates who challenge incumbents. For these types of measures, the excluded groups of candidates are those without political experience (in the case of state legislator focused measures) or those with little chance of winning the election (in the case of donation-based measures). In both instances, candidates without prior political experience and with little chance of winning are most likely to challenge incumbents in the primary stage (Porter and Treul 2023). For context, CFscores, which provide some of the highest levels of coverage of primary candidates among existing measures, did not have a score for 59% of candidates who challenged an incumbent candidate in 2018 and 2020.

In addition to coverage limitations, measures that use approximations for candidate positioning, such as donation behavior, fail to adequately capture changes in candidate po-

[1] It should be noted that approaches using candidates' Tweets could get around this limitation but existing research (e.g., Gaynor et al. 2022) has only collected data on subsets of candidates. Furthermore, a specific limitation of Twitter data is it is shaped by current events that could influence the measurement of candidates positioning from year to year. While there is value in this type of temporal data (for example, see Macdonald et al. 2022), it does not provide a stable picture of aggregate candidate positioning.

sitioning during the campaign. To start, it is likely that citizen behavior is affected by more than just candidates' issue positioning. As Bonica (2014) notes, one example of this are endorsements that can sway donors' perceptions of candidates. But endorsements can happen for reasons unrelated to a candidates' positioning, such as likelihood of winning the election. Further, positioning scores based on donor behavior rely on the assumption that donors actually respond to updated issue positioning rather than relying on previous issue positions and a candidate's reputation. It could also be the case that other electoral factors shape donors' behavior outside the positioning of candidates (Hirano and Snyder 2019). For example, assuming that donors give to the candidate closest to their own position in a congressional primary, the emergence of an extreme (moderate) challenger could siphon off donors from the incumbent who are more extreme (moderate) than the midpoint between the two candidates. If this were to occur, a scaling procedure based on donations would actually cause the incumbent to appear more moderate (extreme) absent this challenger, even if the incumbent did not change her issue positioning.

Data Description

To improve upon the limitations of current measurement approaches, I propose campaign website issue positions as an alternative data source for estimating candidate positioning. Websites are an important part of the modern candidate's campaign. Most candidates in recent years (88% between 2018-2022) maintain a website that acts as an "information hub" for all parts of the campaign, from information about the candidate to their issue positions and policy proposals (Herrnson, Panagopoulos and Bailey 2019). Candidates carefully craft these websites, knowing that potential voters, donors, journalists, and other electoral stakeholders will visit them for information about their campaign (Druckman, Kifer and Parkin 2009). These websites come directly from their campaign, cover a range of issues and policy areas, and are representative of the population of campaigns (Druckman, Kifer and Parkin 2009). Further, throughout an election cycle, little changes on the campaign website, making it a static representation of a campaign for that election (Porter, McDonald and Treul 2021). To this extent, campaign websites are a comprehensive data source for studying candidates in U.S. congressional elections.

As part of their campaign website, most (77%) candidates maintain an "issue page" that explicitly lays out the candidate's stance on the issues, specific policy proposals, and oftentimes commentary on contemporary events. Porter, Treul and Case (2023) collect the issue pages for all Democratic and Republican primary candidates for U.S. House of Representatives who had an official campaign website in 2018, 2020, and 2022. To collect official campaign website issue positions data, Porter, Treul and Case (2023) first identified the names of all candidates running in the primaries from state election boards as well as candidate filings with the Federal Election Commission (FEC). Using this list of names, official campaign websites were identified using a few different sources. `Politics1.com` maintains a database of all campaign websites for candidates running actively in each race and is where the links to a majority of the campaign websites were found. Others were found through various social media pages, `Ballotpedia.com`, and Google searches.

As a part of this data-collection process, research assistants identified whether or not each candidate had a "platform," or a set of issue statements.[2] While this looks different on some websites, it oftentimes is referred to as "My Platform," "Issues," or "Where I Stand." On these issue pages, candidates typically organize their issue stances in a series of paragraphs about different policy areas, or individual issue statements. Research assistants manually collected each of these individual issue statements. This process was done contemporaneously in the ten days leading up to candidates' primary election date, both to ensure consistency in the data collection process and that candidates' websites were finalized in the lead up to the election. In total, this data set contains 4,554 issue pages (77% of all candidates; 87% of candidates with a website). Across 2018, 2020, and 2022, each candidate had a mean of 9.8 individual issue statements on their campaign platform for a total of 44,544 individual issue statements. For a full discussion of the data and the data collection process, see Porter,

[2]Screen shots of example campaign platforms can be found in Appendix A.

McDonald and Treul (2021).[3]

Campaign issue pages improve upon current measurement approaches through both the expansion of the number of candidates included and through actually capturing the issues that candidates run on. To compare the coverage of candidates with an issue page versus previous measurement approaches, Figure 1 plots the number of candidates with an issue page versus the number of candidates with a CFscore for the 2018 and 2020 U.S. House of Representatives primaries.[4] This is further broken down by candidate type: incumbents, those who have previously held elected office, and those who have not previously held elected office. In the aggregate, 2,941 (75%) candidates had an issue page on their campaign website in 2018 and 2020 and 2,522 (65%) have a CFscore.[5] As is evident in Figure 1, campaign websites provide a large increase in coverage of candidates when it comes to inexperienced challengers. Of the 2,654 inexperienced candidates who ran in 2018 and 2020, 1,856 (70%) had an issue page on their campaign website, while only 1,366 (52%) received enough eligible contributions for a CFscore. When it comes to experienced challengers, both sets of data have a high percentage of candidates, with 367 (76%) having an issue page and 392 (82%) out of 483 total experienced challengers having a CFscore in 2018 and 2020. Importantly, it should be noted that a small number of incumbents do not have an issue page on their campaign website, leading to slightly worse coverage with campaign websites (94%) than CFscores (100%).[6]

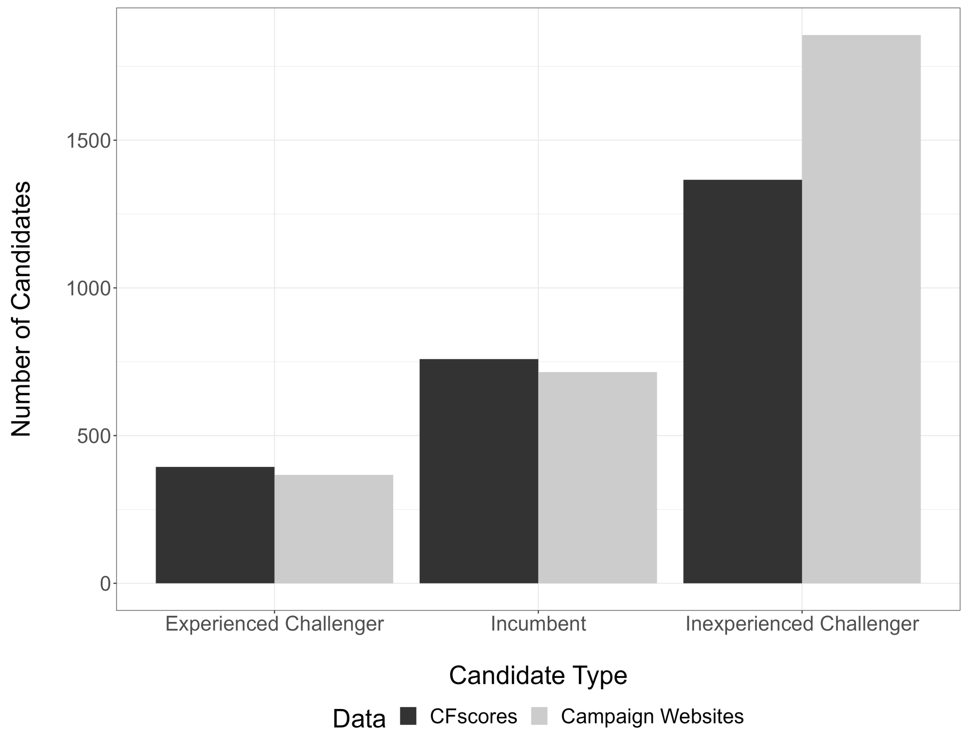
[3]Interviews with campaign consultants who work with candidates on setting up their website highlighted the importance of these pages, mentioning issue pages as the part of the campaign they spent the most time discussing with candidates. It should be noted that while these campaign consultants often use similar strategies across campaigns (Nyhan and Montgomery 2015), interviews highlighted a few important components that ensure the website is capturing candidate behavior. First, while campaign consultants help with the drafting process of issue pages, it is still what the candidate is interested in and wants to focus on for the election that shapes the issue pages. Second, candidates are still the ones operating their campaign, and even with the direction of campaign consultants, the candidate is the one with the final say. Third, despite consistent strategies across the same consulting firms, most have a review process across candidates to ensure that issue text for one candidate is not the same as issue text from another candidate at the same firm; most of the time this process involves separate writers for the issue pages and a secondary check of all issue text. In this manner, these issue pages are individual to each candidate.

[4]2018 and 2020 are used for Figure 1 because these are the current years for which CFscores are available concurrent with the website data collection. Patterns for campaign websites are consistent in 2022.

[5]CFscores are used as a comparison measurement due to the high level of coverage compared with other measures of candidate positioning. Other measures of candidate positioning have a substantially lower percentage of candidates included.

[6]While the number of incumbents without an issue page is a limitation of the measurement, it should

Figure 1: Candidate Coverage by Measurement and Candidate Type



Note: Figure 1 depicts the number of candidates running as either a Democrat or Republican in 2018 and 2020 congressional primary elections who have a CFscore (black bar) and an issue page on their campaign website (gray bar). The data are broken down by challengers with previous elected experience (left), incumbent candidates (middle), and challengers with no previous elected experience.

Moreover, campaign websites do well capturing candidates who did not receive enough eligible contributions for a CFscore. In 2018 and 2020, across all categories of candidates, there were 1,386 candidates without a CFscore. Among those, 742 (54%) have a campaign website issue page. For research that is substantively interested measuring the behavior of primary challengers, or other related work focused on the behavior of inexperienced candidates (see Treul and Hansen 2023), using campaign websites as a data source represents a substantial improvement and improvement over existing measures.

be noted that the number of incumbents with websites and issue pages has increased in the most recent election cycles. This is also true for across other candidate backgrounds. In 2022, coverage of candidates is 366 (97.3%) for incumbents, 225 (79.5%) for experienced challengers, and 1,022 (73.7%) for inexperienced challengers.

Estimation Strategy

To estimate Website EmBedding (WEB) Strategic Positioning Scores using issue position text, I rely on a word embedding model with document-level vectors (Doc2Vec; for the original model specification, see Le and Mikolov 2014) for each candidate-year. Word embeddings are the parameter estimates from neural network models designed to predict word(s) given the context around that word(s). Work in computer science has highlighted the different ways in which word embeddings can capture important underlying properties of language, such as the similarity between words, analogies, and antonyms (Mikolov, Yih and Zweig 2013). Word embedding models have recently seen more wide-spread use in a political science context (Rodriguez and Spirling 2022). Their rise in use stems from the ability to assess and test hypotheses for how word use can differ across covariates (Rodriguez, Spirling and Stewart 2021) as well as uncover important latent traits related to the properties of both words (Grand et al. 2022) and the people using them (Rheault and Cochrane 2020). Moreover, Rodriguez and Spirling (2022) show that word embedding models are able to identify nearest neighbors to politically relevant terms, such as immigration, at the same level as human coders. This suggests embeddings are well-suited to pick up on important semantic relationships in text related to political phenomenon such as candidate positioning.

There are a number of other approaches, both supervised and unsupervised, to estimate candidate or party positioning from text. Underlying both approaches is the assumption that word usage is related to the aggregate positions that candidates take (Grimmer and Stewart 2013). One of the earliest supervised approaches, WordScores, uses a smaller sample of labeled documents, where each document has been labeled by experts to identify their positional leaning. Based on the occurrence of each word in the labeled documents, words then receive a score representing their positional lean. From there, unlabeled documents can then receive a placement based on the occurrence of words and the scores for each word from the previous. However, these supervised methods often conflate positioning reflected in text with stylistic differences in text (Grimmer and Stewart 2013).

Among unsupervised methods, WordFish (Slapin and Proksch 2008) uses regressions to project counts for each word onto each party-year combination. More recently, Vafa, Naidu and Blei (2020) develop text-based ideal points (TBIP) that also uncover specific topics associated with each latent score, providing more validity and taking into account the co-occurrence of words. While WordFish and TBIP improve upon supervised methods by reducing the time and cost of labeling documents, both methods still rely on the occurrence (or co-occurrence in the case of TBIP) of words in a document without taking into account the context of word usage. This contributes to these models having little sense about the semantic relationship between words after the model is estimated (Le and Mikolov 2014). This is an important point when estimating candidate positioning from text. For example, take the words "immigration" and "immigrants." While different parts of speech, both words are semantically similar. An estimation strategy strictly relying on the occurrence of words is not able to account for the semantic similarity between these words. Word embedding models improve upon this limitation in their ability to incorporate high-quality semantic relationships between words during the training process.

The word embedding model I estimate has two parts. The first part of the model is the same as a traditional skip-gram model architecture: a target word, w_t, is used to individually predict the set of words, w_Δ, occurring Δ places before and after w_t in the text. This process repeats over each word in the corpus and word embeddings, the parameter weights in the model, are gradually trained to maximize the ability of the model to predict the words in close proximity to the target word. This ensures that the resulting word embeddings are high-quality representations and capture semantic relationships between words. The second part of the model trains a document vector for each candidate-year. This model architecture is the same as the first part, but instead of using a word embedding to predict words, the candidate embedding replaces the word embedding for the target word and is used to predict the words in w_Δ. Intuitively, this means candidate embeddings are trained to have parameter weights that reflect the word embeddings in candidates' issue statements. In the training process, these two steps are carried out sequentially. As such, this method follows the same assumption of existing approaches that stipulate word usage reflects the aggregate positions of candidates. However, unlike previous approaches, word embeddings account for context (Rheault and Cochrane 2020).

From the model output, each word and candidate-year has an embedding of 300 dimensions. While these embeddings represent a rich understanding of the syntactic and semantic relationship between words and candidates, higher dimension representations are unwieldy for regression analysis. To produce the resulting WEB Scores, I follow Rheault and Cochrane (2020) and use principal component analysis to reduce the candidate embeddings. In determining the number of dimensions, I identify the knee point using the Kneedle algorithm (Satopaa et al. 2011). The algorithm identifies a single dimension as the inflection point resulting in a WEB Score for each candidate in each election. For a full technical explanation of the model, as well as robustness checks relating to the model architecture and parameters, see Appendix B

Measurement Validity

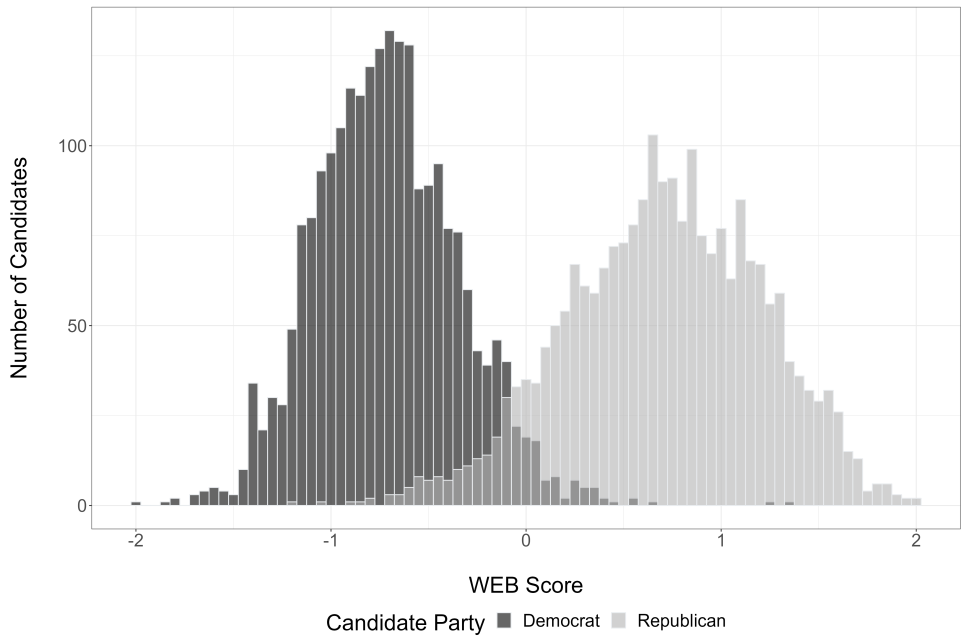
In this section, I provide an overview of the measurement and several validation procedures. The distribution of WEB Scores is plotted in Figure 2. The measurement has a mean of 0 across candidates and a standard deviation of 0.83. Democratic candidates trend to the negative side of the scale with a mean score of -0.71 while Republicans have a mean score of 0.71.[7]

To provide context to the measurement, Table 1 presents the ten most liberal and conservative incumbent candidates for the U.S. House of Representatives from 2018-2022. Notable candidates such as Alexandria Ocasio-Cortez (2018) sit well to the left of the Democratic mean, with a score of -1.31. On the Republican side, Marjorie Taylor Greene (2020) also has a score well to the extreme of the Republican mean at 1.66.

In the sections to follow, I carry out a variety of tests to evaluate the validity of the measurement. It is important to note, when estimating candidate positioning from text in

[7]From the model output, the axis is flipped to follow standard convention with Democrats on the left and Republicans on the right.

Figure 2: Histogram of WEB Scores in 2018, 2020, and 2022 U.S. House of Representatives Primary Elections



Note: Figure shows the distribution of candidates by WEB Score. Democratic candidates are colored dark gray and Republican candidates are colored light gray. Negative values represent more liberal positioning scores and positive values represent more conservative positioning scores.

the American context, there are two important considerations. First – does the measure actually pick up on intra-party differences (Tausanovitch and Warshaw 2017)? Second – does the measure actually text that is related to candidate positioning and not other stylistic choices in the language candidates use (Grimmer and Stewart 2013)? With a focus on both of these hurdles, I first assess the external validity by comparing the measurement with other measures capturing facets of positioning (DW-NOMINATE and CFscores) and show the measures are highly correlated. I also show the measure picks up on various intra-party differences by comparing scoring across various ideological caucuses, a tool members of Congress use to convey their position within parties to voters (Clarke 2020). In addition, I take advantage of the properties of embeddings to show that candidate positioning scores pick up important semantic differences that are related to the positions candidates can take on issues and are reflective of the scale endpoints.

Table 1: Most Liberal and Conservative Incumbent Candidates

Most Liberal	Most Conservative
Barbera Lee (2020)	Troy Bladerson (2020)
Barbera Lee (2022)	Jody Hice (2020)
Nydia Velazquez (2022)	Neal Dunn (2020)
Yvette Clarke (2022)	Neal Dunn (2022)
Nydia Velazquez (2020)	Jeff Duncan (2020)
Melanie Stansbury (2022)	Neal Dunn (2018)
Suzanne Bonamici (2022)	Greg Murphy (2020)
Marilyn Strickland (2020)	William Timmons (2022)
Earl Blumenauer (2020)	Troy Balderson (2018)
Kathy Manning (2022)	Jeff Duncan (2022)

Note: Table shows the most liberal and conservative incumbent candidates who ran in the 2018, 2020, and 2022 congressional elections.

External Validity

To assess the external validity of the WEB Scores, I first evaluate similarity between WEB Scores and pre-existing scores of positioning: CFscores, which capture donors' perceptions of candidates' positioning, and DW-NOMINATE, which captures voting preferences on the congressional legislative agenda. It should be noted that while these concepts are distinct from explicit candidate positioning, they should nonetheless be related. I focus these validation tests on 2018 and 2020 candidates, the election years with both CFscores and DW-NOMINATE.

Table 2 shows the correlations (standard errors in parentheses) for candidates elected to Congress in 2018 and 2020, restricted to those having a WEB Score, a CFscore, and a DW-Nominate score in the 116th and 117th Congress.[8] The first panel looks at all candidates, the second looks at Democratic candidates, and the third looks at Republican candidates. Starting with all candidates, the correlation between WEB Scores and DW-Nominate is high at 0.88, as well as the correlation between WEB Scores and CFscores at 0.89. These correlations are substantively similar to the correlation between DW-Nominate and CFscores at 0.93.

[8]It should be noted this restricts the comparisons to only candidates who were elected to Congress. When looking at all candidates with a CFScore and a WEB Score, the correlations are similar: the correlation for all candidates is 0.86 (0.0001), Democratic candidates is 0.25 (0.0007), and Republican candidates is 0.21 (0.0011).

Table 2: Measure Correlations for 116th and 117th Congress

All Members of Congress			
	CFscores	DW-NOMINATE	WEB Scores
CFscores	1.00	–	–
DW-NOMINATE	0.93 (0.0002)	1.00	
WEB Scores	0.89 (0.0003)	0.88 (0.0003)	1.00
Democrats			
	CFscores	DW-NOMINATE	WEB Scores
CFscores	1.00	–	–
DW-NOMINATE	0.03 (0.0023)	1.00	–
WEB Scores	0.43 (0.0019)	0.21 (0.0023)	1.00
Republicans			
	CFscores	DW-NOMINATE	WEB Scores
CFscores	1.00	–	–
DW-NOMINATE	0.54 (0.0019)	1.00	–
WEB Scores	0.28 (0.0025)	0.48 (0.0021)	1.00

Note: Table 2 shows the correlation coefficient (standard error in parentheses) between CFscores, DW-NOMINATE, and WEB Scores for candidates running in 2018 and 2020 who have a score for all three measures. The first panel includes candidates from both parties, the second panel includes only Democratic candidates, and the third panel includes only Republican candidates. When looking at all candidates with a CFScore and a WEB Score, the correlations are similar: the correlation for all candidates is 0.86 (0.0001), Democratic candidates is 0.25 (0.0007), and Republican candidates is 0.21 (0.0011).

Turning to intra-party correlations for Democratic candidates, WEB Scores are weakly correlated with DW-Nominate at 0.21 but moderately correlated with CFscores at 0.43.[9] Both are significantly higher than the correlation between CFscores and DW-NOMINATE for Democrats (0.03). Among Republican candidates who were elected to Congress, the correlation between WEB Scores and DW-NOMINATE is moderate at 0.48. This is, again, similar to the intra-party correlations for CFscores and DW-Nominate at 0.53. The correlation between WEB Scores and CFscores is substantively lower at 0.28.

In the aggregate, these correlations provide evidence the measures are capturing related but distinct concepts, as expected. But the intra-party correlations highlight varying levels of validity when comparing WEB Scores with existing measures within party. To further validate the measure against previously established proxies for candidate positioning, I also

[9]It should be noted that part of this can likely be attributed to shortcomings of DW-NOMINATE in correctly identifying more liberal members of Congress. Correlations with alternative estimation strategies, such as Duck-Mayr and Montgomery (2022), would likely be higher within the Democratic Party given this adjustment.

compare the average WEB Score for each ideological caucus in Congress. Members view ideological caucuses as a means to convey their positioning to donors and voters, especially within parties (Clarke 2020). Considering that joining an ideological caucus can be motivated by electoral interests, WEB scores should reflect positioning differences among caucus groups *within* parties.

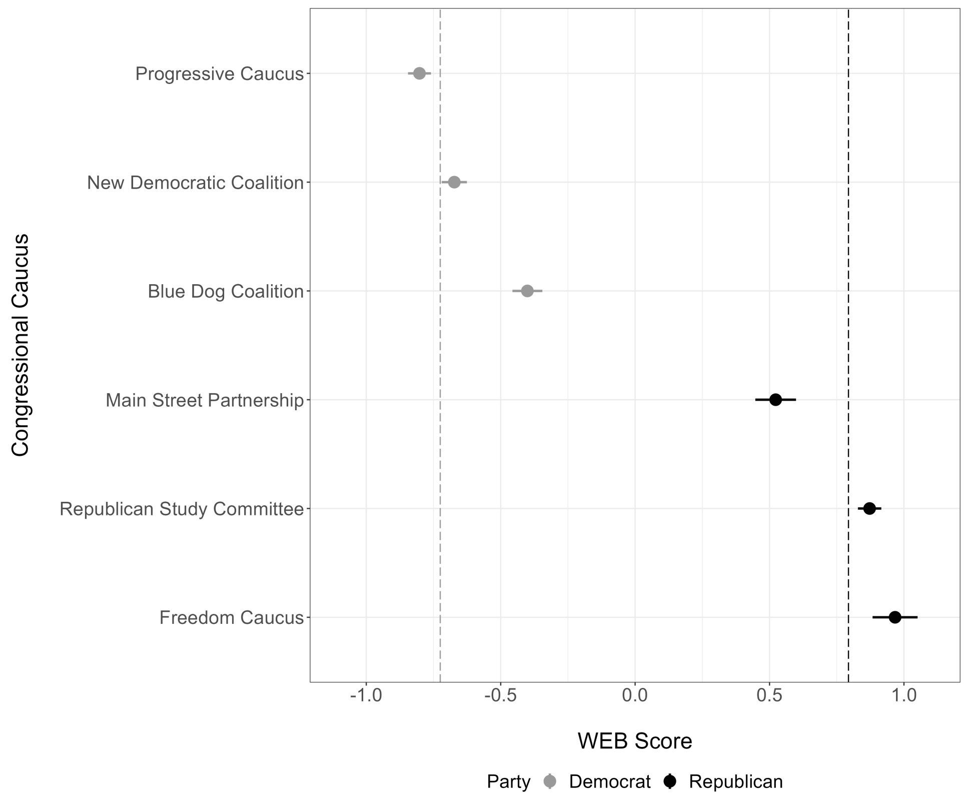
To do this, I collect ideological caucus membership data for six caucuses that are relevant for the 2018, 2020, and 2022 elections (from most liberal to most conservative, according to Clarke (2020)): the Progressive Caucus, the New Democratic Coalition, the Blue Dog Coalition, the Main Street Partnership, the Republican Study Committee, and the Freedom Caucus. Caucus membership is collected for incumbent candidates during the Congress that runs concurrently with the election (e.g., incumbent candidates running in the 2018 election and caucus membership in the 115th Congress running from 2017 to 2019). Data for the 115th Congress comes from Clarke (2020) while the 116th and 117th Congresses were collected from archived official caucus membership pages. Because the Freedom Caucus does not maintain an official caucus roster online, the membership was gathered from a news article[10] that provided a roster of Freedom Caucus members.[11]

Figure 3 plots the mean WEB Score by caucus for incumbent candidates running in 2018, 2020, and 2022, as well as 95% confidence intervals. In addition, vertical dashed lines depict the mean WEB Scores for incumbent candidates in the Democratic and Republican parties, respectively. Starting with the Democratic Party, WEB Scores pick up on intra-party differences by caucus membership. Incumbent candidates in the Progressive Caucus have the lowest score at an average of -0.80. This is less than the New Democratic Coalition, which has a mean of -0.67 (diff $= -0.13$, p-value ≤ 0.001). The New Democratic Coalition has an average value significantly larger than the Blue Dog Coalition, which has a mean WEB Score of -0.40 (diff $= -0.27$, p-value ≤ 0.001). The differences and the ideological caucus ordering match expectations and provide face validity the measurement is picking up

[10]Newsweek, "Who Is In House Freedom Caucus? Full List of Members After Midterms Results" November 10, 2022

[11]Caucus membership for Freedom Caucus members who served across multiple Congresses is consistent with Clarke (2020).

Figure 3: Average WEB Score by Caucus Membership



Note: Figure plots the mean WEB Score for incumbent candidates by ideological caucus with 95% confidence intervals. Ideological caucuses are ordered on the y-axis from liberal (top) to conservative (bottom) according to (Clarke 2020). The average position scores of all ideological caucuses are statistically different from one another at the p<0.05 level.

on intra-party differences within the Democratic Party.

Turning to Republican incumbent candidates, WEB Scores also pick up on expected differences by caucus within the party. The more moderate Main Street Partnership has a mean of 0.52. Both the Republican Study Committee, with a mean of 0.87 (diff = 0.25, p-value $\leq$ 0.001), and the Freedom Caucus, with a mean of 0.97 (diff = 0.45, p-value $\leq$ 0.001), have average WEB Scores greater than the Main Street Partnership. WEB Scores also pick up on differences between the Republican Study Committee and the Freedom Caucus, with the Freedom Caucus having a higher average value (diff = 0.10, p-value $\leq$ 0.05). Within both parties, the differences in mean caucus scores provide face validity the measure picks up on differences in candidate positioning within parties.

Internal Validity

In addition to validating WEB Scores relationship with other related constructs, I also test whether or not WEB Scores capture underlying constructs in the actual text that is related to candidate positioning. One of the advantages of word embedding models is the ability to uncover semantic relationships between words using arithmetic, sometimes referred to as linear substructures. In the classic example from Mikolov et al. (2013), the authors are able to show:

$$vector[\text{“}king\text{”}] - vector[\text{“}man\text{”}] + vector[\text{“}women\text{”}] = vector[\text{“}queen\text{”}]$$

The ability to uncover these types of semantic relationships between words makes it possible to test a variety of word relationships that should be related to candidate positioning, thus validating the measure against the underlying text. This is possible because word embeddings and candidate embeddings exist in the same dimensional space. As such, if WEB Scores are capturing variation in candidate positioning, they should also be related to certain semantic relationships that capture various positions. Take the following example between the word "universal" and the word "healthcare." Given that advocating for universal healthcare is more often done by liberal candidates, it should be expected that the relationship between these words is closer for liberal candidates (e.g., Alexandria Ocasio-Cortez (D-NY)) than with conservative candidates (e.g., Chip Roy (R-TX)). This comparison can be done by adding the candidate embedding to the word embedding for healthcare and then assessing the cosine similarity between the new embedding and the word embedding for "universal."[12] It should be expected this similarity is greater for the more liberal candidate. Specifically:

$$\begin{aligned} &cosine(vector[\text{“}healthcare\text{”}] + vector[\text{“}Ocasio-Cortez2022\text{”}], vector[\text{“}universal\text{”}]) \geq \\ &\qquad cosine(vector[\text{“}healthcare\text{”}] + vector[\text{“}Roy2022\text{”}], vector[\text{“}universal\text{”}]) \end{aligned}$$

[12] Cosine similarity assesses the angle between the two vectors. This method is ideal for capturing vector similarities in a high number of dimensions.

As expected, the cosine similarity for Ocasio-Cortez is 0.31 versus 0.15 for Roy, showing the semantic similarity between "universal" and "healthcare" is closer for Ocasio-Cortez than it is for Roy.

To more formally carry out this test, I rely on the notion that $vector[candidate] + vector[policy]$ should be more similar to a conservative (liberal) policy proposal embedding across candidates as WEB Scores increase (decrease). In developing policy proposal embeddings, I rely on Distributed Dictionary Representations (DDR; Garten et al. 2018). The advantage of this method is that by averaging word embeddings, it is possible to capture a distinct psychological construct. For the purposes of this paper, I use DDR to develop average policy position embeddings that can conceivably be classified as either more liberal or more conservative. To do so, I rely on a set of eight anchoring vignettes that represent the end points of the position scales – four from the Justice Democrats Policy Priorities in 2022, and four from the Heritage Foundations Policy Priorities in 2022.[13]

From each of the policy priorities, I select a set of keywords that are present in the stance each organization is taking. The policy word, policy stance, and policy proposal keywords can be found in Table 3. Full issue statement vignettes can be found in Appendix C.[14] To carry out the test, I add each candidate embedding to the word embedding for each policy area (e.g., government). I calculate an average embedding of the keywords, and calculate cosine similarities between the resulting candidate-policy embedding and the keyword embedding for each candidate in each policy area.

After calculating the relevant cosine similarities, I conduct eight OLS regressions where the dependent variable is the cosine similarity for each policy area and the independent

[13]The Justice Democrats and Heritage Foundation are chosen because they lay out clear, detailed policy positions and self-describe as placing themselves at the extreme of the positioning scale. Justice Democrats outline their mission "is to build a mission-driven caucus in Congress by electing more leaders like Alexandria Ocasio-Cortez and Jamaal Bowman, who will represent our communities in Congress and fight for bold, progressive solutions to our current crises." The Heritage Foundation states their mission is to "formulate and promote public policies based on the principles of free enterprise, limited government, individual freedom, traditional American values, and a strong national defense." This provides face validity to the anchoring vignettes.

[14]One of the advantages of embeddings, and DDR specifically, is that not all words need to be included in the dictionary. For example, because regulation and regulations are syntactic pairs, the inclusion of both adds little to the set of keywords.

Table 3: Policy Word, Policy Stances, and Keywords for Internal Validity Test

Policy Area	**Policy Stance**	**Keywords**
Abortion (Heritage)	Banning abortions after fetal heartbeat	prolife, families, heartbeat, prohibit
Education (Heritage)	Increasing parental involvement in curriculum	parents, choice, homeschooling, transparency
Government (Heritage)	Reducing government spending and regulation	spending, regulations, prudent, fiscal
Immigration (Heritage)	Increasing border security	incursions, enforces, prosecutes, secures
Environment (Justice Democrats)	Increasing renewable energy sources and protecting vulnerable communities	renewable, climate, fossil, color
Guns (Justice Democrats)	Increasing gun control	background, ban, assault, safety
Healthcare (Justice Democrats)	Implementing single-payer health insurance	universal, singlepayer, expand, medicareforall
Wages (Justice Democrats)	Increasing the minimum wage	living, minimum, affordable, cost

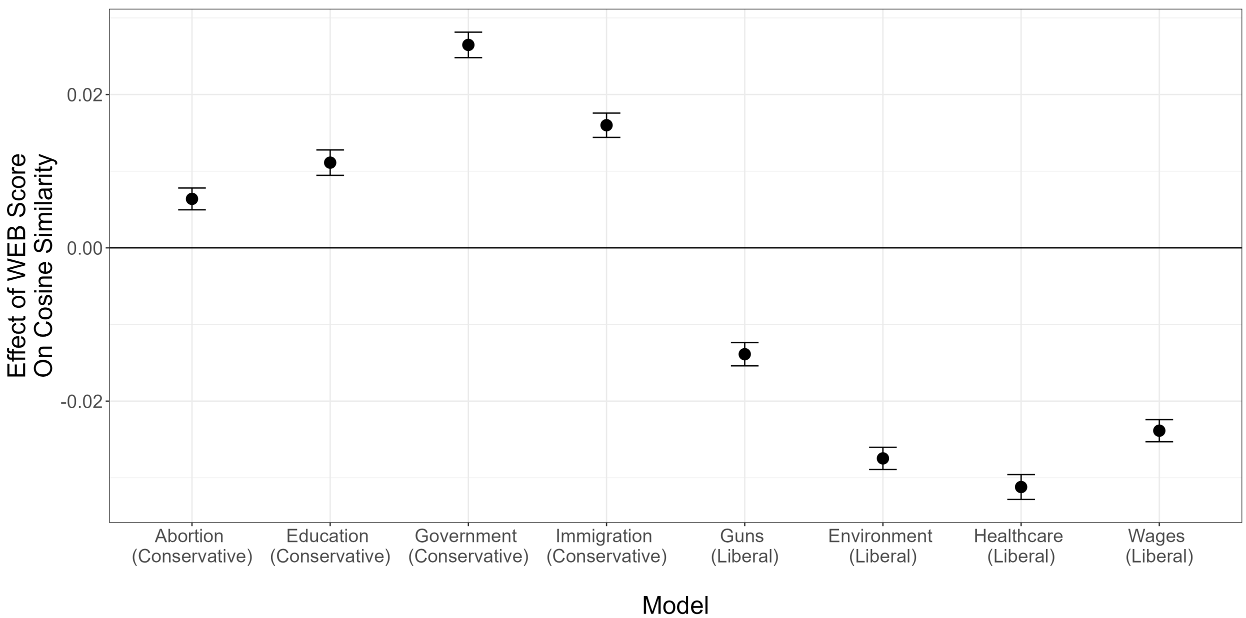
Note: Table displays the policy area (left column), the associated policy proposal advocated for by either the Justice Democrats (liberal) or the Heritage Foundation (conservative; middle column), and the keywords used in the policy proposal (right column). Full text related to policy proposal can be found in Appendix C.

variable is candidates' WEB Score. If WEB scores are picking up on important semantic relationships related to candidate positioning, it should be expected that the coefficient is positive (negative) for conservative (liberal) policies.

Figure 4 plots the coefficient for WEB Scores from all eight models.[15] Conservative policies are on the left side of the table and liberal policies are on the right side of the table. Across the four conservative policies, the effect of WEB Scores is positive. This means that as WEB Scores increase, the cosine similarity between $vector[candidate] + vector[policy]$ and the average of $vector[keywords]$ for each policy area increase. This can be interpreted as the word embedding for policy words and the average word embedding for the keywords as being more similar for conservative candidates than liberal candidates. For the liberal policies, the effect is negative, as expected. This means as WEB Scores decrease, the cosine similarity between $vector[candidate] + vector[policy]$ and the average of $vector[keywords]$ for each policy area increase. These results provide evidence the measure is picking up on

[15] Full regression tables can be found in Appendix D.

Figure 4: Effect of Candidate WEB Scores on Policy Cosine Similarities



Note: Figure presents coefficient estimates and 95% confidence intervals from the effect of WEB Scores on cosine similarities for each candidate and the relevant policy area. Full regression output can be found in Appendix D.

positional differences across candidates in text, further validating the resulting measurement.

Analysis

Given the validity of WEB Scores, as well as their advantages as a measure of campaign positioning over existing measures, they are well situated to assess whether or not incumbents respond to the positioning of primary challengers during the campaign. In all models, the dependent variable of interest, incumbent position extremity, is measured as an incumbent candidate's WEB Score minus their party's average WEB Score. I then multiply the Democratic candidates' score by -1 to provide a consistent measure across parties (Hirano and Snyder 2019). Therefore, positive (negative) values are interpreted as candidates becoming more extreme (moderate).

For the key independent variable of interest, the positioning of a primary challenger, I focus on only the challenger with the highest vote share. This is done because in many races with more than one challenger, the candidate with the highest vote share is likely the most serious and the one an incumbent would be most likely to respond to.[16] To test this theory, I rely on two measurement approaches for the independent variable. The first approach is dichotomous and takes on the values of "Extreme Challenger," "Moderate Challenger," or "No Challenger (Reference Category)." I classify challengers as extreme if they had a WEB Score greater than their parties' mean WEB Score, and moderate otherwise.[17] For the continuous approach, I measure the challenger positioning in the same way I measure the dependent variable: I subtract the challenger's party mean WEB Score from the challenger's own WEB Score. I then multiply Democratic candidates' measure by -1 for a consistent measure across parties. It should be noted, that in the continuous approach, because challenger positioning is not observed when a challenger does not emerge, this model only includes candidates who are challenged in the primary. The dichotomous approach includes all incumbents.

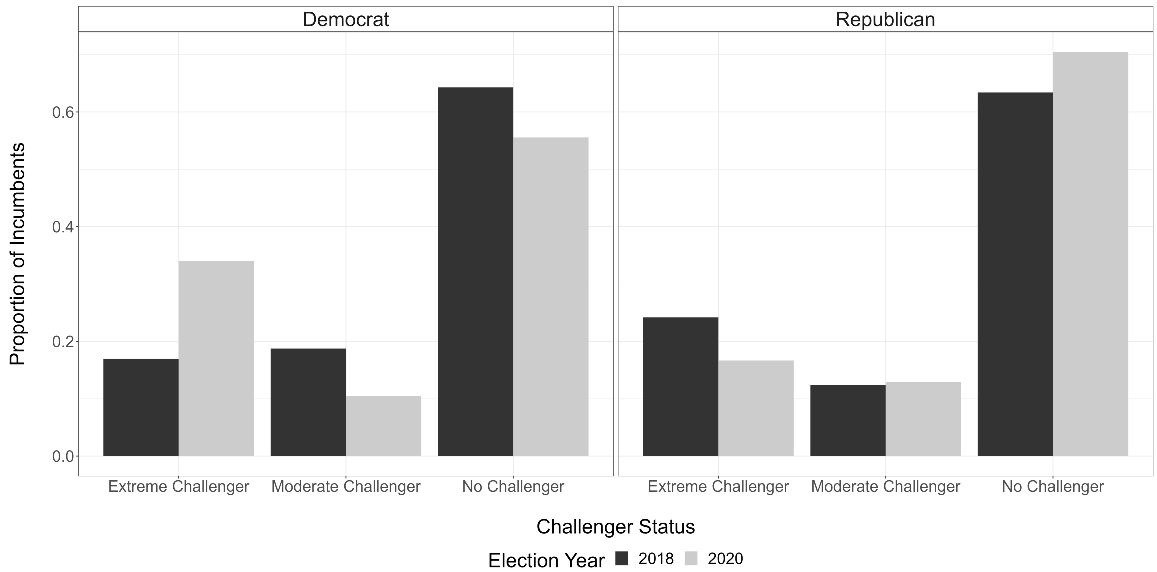
To provide context for where incumbents face primary challenges from, Figure 5 plots the percent of incumbents who were either challenged by an extreme candidate, a moderate candidate, or were not challenged by party and election year.[18] Consistent with prior research, the majority of incumbents in both parties and election years are able to stave off primary challengers (63%), although there is a slight divergence in the trend across parties from 2018 to 2020 – more Democratic incumbents were challenged in 2020 (56%) than 2018 (64%) while fewer Republican incumbents were challenger in 2020 (70%) than 2018 (64%). This likely reflects the electoral environment and circumstances surrounding the 2020 election that presented favorable electoral circumstances to Democrats.

[16] It should be noted, this decision only applies in cases where there are more than one challenger, which constitutes 19% of races within the scope of analysis. Of those races with more than one challenger, only 25% are instances where challenger positioning would be coded differently depending on the challenger, or 8% of all observations. Further, Appendix E shows the results are not sensitive to either (1) changing the direction of the coding in these 8% of races (see Table 7 in the appendix) or (2) coding "Extreme Challenger" and "Moderate Challenger" as equal to 1 in cases where incumbent candidates are cross-pressured (see Table 6 in the appendix).

[17] An alternative specification would classify candidates as an extreme (moderate) challenger if they were more extreme (moderate) than the incumbents' positioning in the previous election. The lack of WEB Scores for 2016 prevent the pursuit of this approach. It can be noted, however, that in the 2020, 81% of challenging candidates would be classified the same using this approach as the approach in the main results, suggesting the results would be consistent.

[18] Incumbents where the challenger does not have a WEB Score are excluded. Figure 5 only categorizes the challenger with the highest vote share, although descriptive patterns are consistent when expanding to all challengers.

Figure 5: Incumbent Challenger Status by Party and Year



Note: Figure depicts the percent of incumbent candidates by running in partisan primaries for 2018 and 2020 party-year and challenger status in the primary election. Challengers are considered extreme if they have a WEB Score more extreme than their party's mean score and moderate otherwise.

When considering the positioning of the challenger, overall, incumbents more often face a challenger from extreme candidates (24%) than moderate candidates (13%). However, these trends do vary by party and election year. Republican incumbents in 2018 were challenged at a higher rate by extreme candidates (24%) than moderate candidates (12%); there was no discernible difference between these two groups in 2020 (17% versus 13%). For Democrats, there were actually more incumbents facing moderate primary challengers in 2018 (19%) than extreme candidates (17%). In 2020, this trend was drastically different, with 34% of incumbents facing an extreme primary challenger while only 10% faced a moderate primary challenger.

To empirically test how incumbent candidates respond to primary challengers positioning, I include incumbent-level random effects and year fixed effects.[19] Among the 406 candidates who appeared on a partisan primary ballot as an incumbent in either 2018 or 2020, 236 (58%) of those appeared as an incumbent in both 2018 and 2020. Because the majority of incumbents over the period of study run in multiple elections as an incumbent, the random

[19] I focus on models with random effects due to the limited number of observations. Table 8 of Appendix E presents the same models using incumbent fixed effects and the results are consistent across all models.

effects by incumbent account for unobserved heterogeneity that could affect a candidate's positioning. In essence, the random effects by incumbent account for baseline levels of a candidate's positioning that could be affected by district characteristics, primary type, or candidate characteristics such as gender or seniority, among other factors. As a result, the coefficient can be interpreted as capturing within-incumbent variation as a result of changes in the status and positioning of a primary challenger.[20]

In addition to the two measurement approaches using WEB Scores, I carry out the same measurement strategy using CFscores. As discussed previously, there are a few reasons to expect different effects between the two underlying measures. For one, CFscores exclude a large number of candidates, especially those likely to challenge an incumbent. For context, in 2018 and 2020, 295 incumbents faced a challenger in the primary. Among those, 31% do not have a WEB Score while 59% do not have a CFscore. While a large proportion of candidates are excluded by CFscores, CFscores still cover more of these candidates than existing measures. Furthermore, given the CFscores approximate positioning using donor behavior, it is possible changes in scores are due to donors changing their behavior in response to changing electoral dynamics, not incumbents. In this case, if donors give to proximate candidates, it is possible the inclusion of an extreme (moderate) challenger actually causes incumbents to appear more moderate (extreme), even if they do not change their behavior.

Table 4 presents the result of both measurement approaches. The first two columns include all incumbents running in partisan primaries in 2018 and 2020 using the dichotomous variable approach specified above. The last two columns include only incumbents who were challenged in a primary election and use the continuous measurement approach. Columns 1 and 3 use WEB Scores, and columns 2 and 4 use CFscores.

Starting with the dichotomous variable approach using WEB Scores in the first column, the results provide support for the hypothesis that incumbent candidates change their behavior in response to primary challengers. When compared with no challenger, incumbents

[20]While I focus on 2018 and 2020 to draw a direct comparison with CFscores, I also run the same models with 2018, 2020, and 2022 for WEB Scores. The results are consistent and can be found in Table 9 of Appendix E.

Table 4: Incumbent Positioning and Challenger Extremity

	Measurement:			
	WEB Scores	CFscores	WEB Scores	CFscores
	(1)	(2)	(3)	(4)
Extreme Challenger	0.081^{**}	-0.017		
ref: No Challenger	(0.037)	(0.020)		
Moderate Challenger	-0.097^{**}	0.038		
ref: No Challenger	(0.044)	(0.038)		
Challenger Extremity			0.169^{***}	-0.003
			(0.057)	(0.036)
Constant	0.020	-0.282^{***}	0.054	-0.357^{***}
	(0.026)	(0.021)	(0.039)	(0.045)
Observations	518	469	189	122
Candidate Random Effects	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓

Note: Units of analysis in the first two columns include all incumbent candidates running in partisan primaries in 2018 and 2020 with a valid positioning score and either (1) a primary challenger who has a valid positioning score or (2) no primary challenger. The second two columns include all incumbent candidates running in partisan primaries in 2018 and 2020 with a valid positioning score and a primary challenger with a valid positioning score. *p<0.1; **p<0.05; ***p<0.01

facing an extreme (moderate) primary challenger position themselves more to the extreme (middle). To consider the substantive implications of this result, the effect of going from no challenger to an extreme challenger is 0.097, or the same difference of means between the Republican Study Committee and the House Freedom Caucus (0.10) from Figure 3. Similarly, the effect of going from a moderate challenger to an extreme challenger is 0.176, larger than the difference between the New Democratic Coalition and the Progressive Caucus (0.13). Both results demonstrate substantively important effect sizes.

Turning to the second column, which follows the same measurement strategy using CF-scores, there is no statistically significant effect on the status of the primary challenger on an incumbent's positioning. Part of this can be attributed to the decline in sample size – 52 incumbents who were challenged in the primary do not have a CFscore for their challenger when compared with WEB Scores – as well as the data generating process of CFscores. Consistent with the expectations laid out above regarding donors responding to changing electoral dynamics, the coefficient is in the opposite direction.[21]

Turning to the third and fourth columns, which look at only incumbents who were challenged in the primary, the results are consistent with the first two columns. When using WEB Score, the extremity of the challenger is associated with an increase in the extremity of the incumbent after accounting for incumbent random effects. As expected, there is discernible effect when using CFscores.

Conclusion

This paper provides an important contribution to our understanding of incumbent positioning in response to primary challengers, as well as a broader understanding of the implications of primary elections on American polarization. While theories of democratic competition, and anecdotal evidence from races such as Rep. Kim's, provide a strong argument for incumbents adopting a more extreme (moderate) position in response to an extreme (moderate) challenger, prior work had mostly failed to find support for this theory. As I argue, this is in large part due to measurement limitations and a focus on legislative behavior; when extending this analysis to a measurement (WEB Scores) that covers a larger scope of primary elections and is actually based on the issues candidates take, there is support for this theory.

In addition to the substantive contribution, I also introduce Website EmBedding (WEB) Strategic Positioning Scores which improves upon the limitations of prior measurements of strategic candidate positioning. Namely, it increases the number of candidates included and is based on underlying data that actually captures candidate positioning. In addition, the measure possesses high construct validity, both with other aspects of candidate positioning

[21] To further demonstrate this is not strictly due to the sample size, I run the models using a consistent sample of incumbents who are not excluded due to missingness in either measure. The results are consistent with Table 4 and are presented in Table 10 of Appendix E.

and with the underlying campaign text. The benefits of this new measure, as well as the word and candidate embeddings, expand the number of substantive research questions that can be answered as it relates to candidate positioning.

An important takeaway from the results of this paper is the implications measurement choice can have on the substantive conclusions researchers draw, particularly as it relates to measuring candidate positioning. While CFscores, DW-NOMINATE, WEB Scores, and other measures of elite positioning are all capturing related concepts, they are distinct and based on different underlying data. It is important when choosing a measure that researchers consider what underlying behavior is actually expected to change, and what measure, whether it is one of the three above or others mentioned in this paper, is best suited to capture that construct. For example, as Tausanovitch and Warshaw (2017) note, very few measures provide more predictive validity than DW-NOMINATE when it comes to legislative behavior; WEB Scores should not be seen as, nor treated as, a measure that is predictive of legislative behavior. It should be noted that there are certain scope conditions where researchers may be forced into one measure over another. For example, WEB Scores currently exist going back to 2018 and do not currently exist for candidates who do not run for office (e.g., state legislators). This should not prevent researchers from asking substantively important questions where the perfect measure does not exist. Rather, it is important future research considers the data generating process and other factors that could be contributing to their results and be forthcoming with these limitations.

References

Ahler, Douglas J, Jack Citrin and Gabriel S Lenz. 2016. "Do Open Primaries Improve Representation? An Experimental Test of California's 2012 Top-Two Primary." *Legislative Studies Quarterly* 41(2):237–268.

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

Ansolabehere, Stephen and Philip Edward Jones. 2010. "Constituents' Responses to Congressional Roll-Call Voting." *American Journal of Political Science* 54(3):583–597.

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

Black, Duncan. 1948. "On the rationale of group decision-making." *Journal of political economy* 56(1):23–34.

Boatright, Robert G. 2014. *Getting Primaried: The Changing Politics of Congressional Primary Chal- lenges*. University of Michigan Press.

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

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.

Christopher, Hare, Bakker Ryan, Carroll Royce et al. 2015. "Using Bayesian Aldrich-McKelvey Scaling to Study Citizens' Ideological Preferences and Perceptions." *American Journal of Political Science* 59(3):759–774.

Clarke, Andrew J. 2020. "Party Sub-Brands and American Party Factions." *American Journal of Political Science* 64(3):452–470.

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

Downs, Anthony. 1957. "An Economic Theory of Political Action in a Democracy." *Journal of Political Economy* 65:135–150.

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. 2022. "Ends against the middle: Measuring latent traits when opposites respond the same way for antithetical reasons." *Political Analysis* pp. 1–20.

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.

Gaynor, SoRelle W, Kristina Miler, Pranav Goel, Alexander M Hoyle and Philip Resnik. 2022. "Do You Walk the Walk, Talk the Talk, or Tweet the Tweet? Legislators' Ideal Points Across Venues.".

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 behaviour* 6(7):975–987.

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. and James M. Snyder. 2015. "Candidate Ideology and Electoral Success.".

Herrnson, Paul S, Costas Panagopoulos and Kendall L Bailey. 2019. *Congressional elections: Campaigning at home and in Washington*. Cq Press.

Hirano, Shigeo, Gabriel S Lenz, Maksim Pinkovskiy and James M Snyder Jr. 2015. "Voter learning in state primary elections." *American Journal of Political Science* 59(1):91–108.

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

Jacobson, Gary C. 2012. "The Electoral Origins of Polarized Politics: Evidence From the 2010 Cooperative Congressional Election Study." *American Behavioral Scientist* 56(12).

Jewitt, Caitlin E and Sarah A Treul. 2019. Ideological primary competition and congressional behavior. In *Congress & the Presidency*. Vol. 46 Taylor & Francis pp. 471–494.

Lau, Jey Han and Timothy Baldwin. 2016. "An empirical evaluation of doc2vec with practical insights into document embedding generation." *arXiv preprint arXiv:1607.05368* .

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

Macdonald, Maggie, Megan A Brown, Joshua A Tucker and Jonathan Nagler. 2022. "To Moderate, Or Not to Moderate: Strategic Domain Sharing by Congressional Campaigns." *Or Not to Moderate: Strategic Domain Sharing by Congressional Campaigns (April 5, 2022)* .

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado and Jeffrey Dean. 2013. "Distributed Representations of Words and Phrases and their Compositionality." *Advances in neural information processing systems* .

Mikolov, Tomáš, Wen-tau Yih and Geoffrey Zweig. 2013. Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies*. pp. 746–751.

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.

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.

Pearson, Kathryn and Jennifer L Lawless. 2008. Primary Competition and Polarization in the US House of Representatives. In *Annual Meeting of the Midwest Political Science Association, Chicago, IL*.

Porter, Rachel, Maura McDonald and Sarah Treul. 2021. "Changing the Dialogue: Descriptive Candidacies and Position Taking in Campaigns for the US House of Representatives.".

Porter, Rachel and Sarah A. Treul. 2023. "Evaluating (In)Experience in Congressional Elections." `https://rachelporter.org/files/inexperience.pdf`.

Porter, Rachel, Sarah A. Treul and Colin R. Case. 2023. "Database on Primary Election Website Content." *Chapel Hill, NC; University of North Carolina Libraries* .

Ramey, Adam. 2016. "Vox populi, vox dei? Crowdsourced ideal point estimation." *The Journal of Politics* 78(1):281–295.

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

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

Rodriguez, Pedro L, Arthur Spirling and Brandon M Stewart. 2021. "Embedding Regression: Models for Context-Specific Description and Inference." *American Political Science Review* pp. 1–20.

Satopaa, Ville, Jeannie Albrecht, David Irwin and Barath Raghavan. 2011. Finding a" kneedle" in a haystack: Detecting knee points in system behavior. In *2011 31st international conference on distributed computing systems workshops*. IEEE pp. 166–171.

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

Sides, John, Chris Tausanovitch, Lynn Vavreck and Christopher Warshaw. 2020. "On the representativeness of primary electorates." *British Journal of Political Science* 50(2):677–685.

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.

Sulkin, Tracy. 2005. *Issue Politics in Congress*. Cambridge University Press.

Sulkin, Tracy. 2009. "Campaign Appeals and Legislative Action." *Journal of Politics* 71(3):1093–1108.

Tausanovitch, Chris and Christopher Warshaw. 2017. "Estimating candidates' political orientation in a polarized congress." *Political Analysis* 25(2):167–187.

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

Thomsen, Danielle M. 2020. "Ideology and gender in US House elections." *Political Behavior* 42:415–442.

Treul, Sarah A and Eric R Hansen. 2023. "Primary Barriers to Working Class Representation." *Political Research Quarterly* p. 10659129231154914.

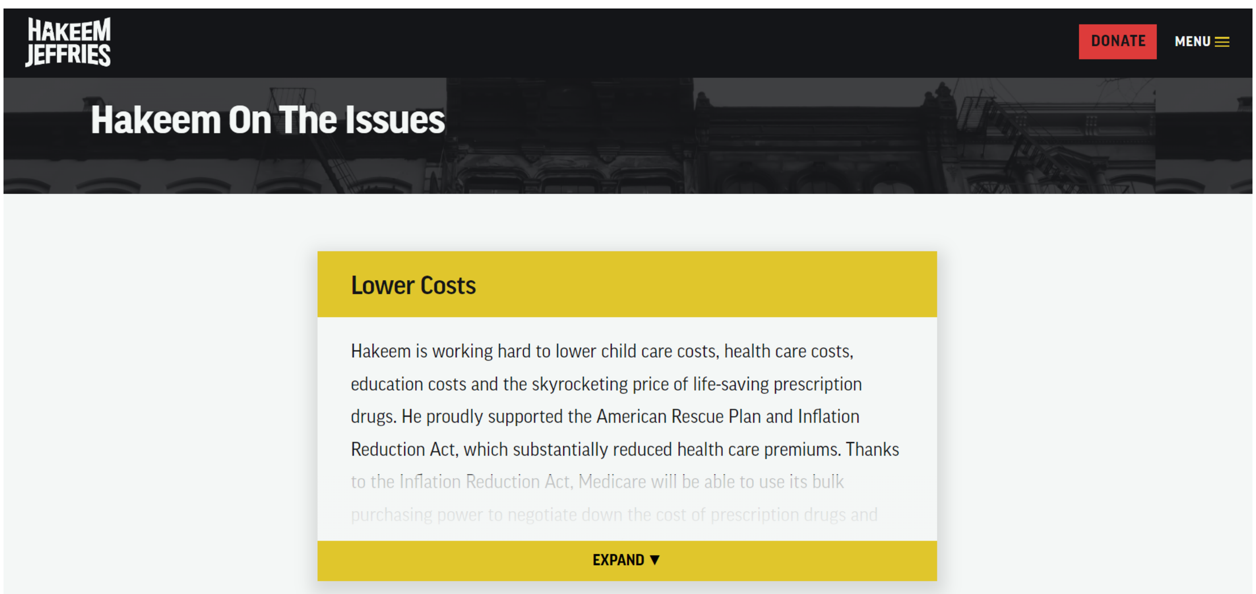
Vafa, Keyon, Suresh Naidu and David M Blei. 2020. "Text-based ideal points." *arXiv preprint arXiv:2005.04232* .

Supporting Information Table of Contents

Appendices

A Example Website Issue Pages

Figure 1: Examples of Campaign Issue Pages from 2022 Congressional Primary Candidates



Note: The top image is from Rep. Dan Crenshaw (R-TX) and the bottom image is from Rep. Hakeem Jeffries (D-NY).

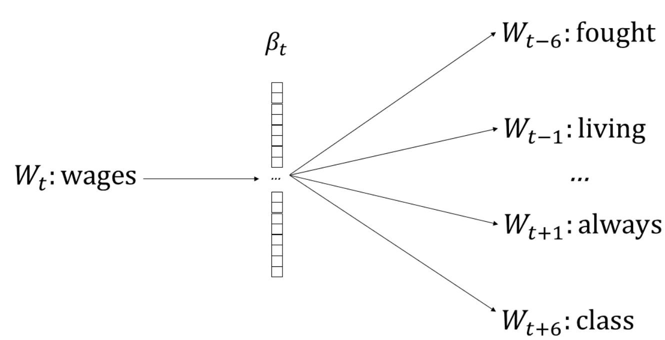
B Word Embeddings Overview and Robustness Checks

The following appendix provides a more detailed description of the word embedding model used to estimate WEB Scores as well robustness checks related to the model architecture and parameters. To estimate WEB Scores, I rely on a word embedding model with document-level vectors for each candidate-year occurrence following the Paragraph Vector Distributed Bag of Words (PV-DBOW) approach developed by (Le and Mikolov 2014). My implementation differs slightly from the original approach. The traditional PV-DBOW implementation does not store word embeddings. While this leads to a more efficient estimation (Le and Mikolov 2014), the quality of the results is inconsistent (Lau and Baldwin 2016). For this reason, I follow Lau and Baldwin (2016) and use a simultaneous skip-gram word embedding model. The following subsections outline the model architecture, implementation, and robustness

Model Architecture

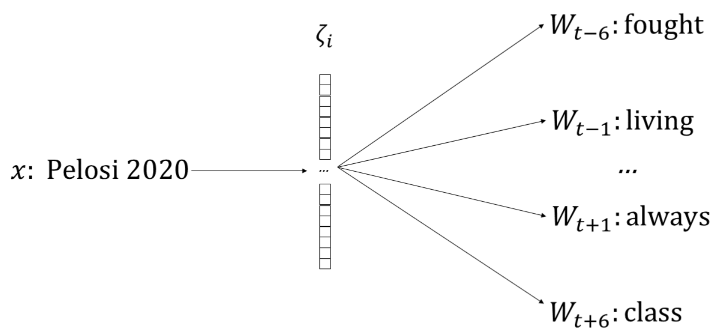
In the model, each candidate-year, i, and word, j, has an embedding with M dimensions, denoted as ζ_i and β_j respectively. The model has two parts. The first part of the model follows a traditional skip-gram model architecture developed by Mikolov et al. (2013). In this estimation, for each document, a word, w_t is sampled at each iteration and the window, Δ, before and after that word is extracted. The resulting window surrounding w_t, denoted as w_Δ, are the outcomes of interest. The output can be written more completely as $w_\Delta = (w_{t-\Delta}, \ldots, w_{t-1}, w_{t+1}, \ldots, w_{t+\Delta})$. The model input is an indicator vector, x_t, for the target word, w_t. x_i is multiplied by the matrix of candidate embeddings, β. The resulting word embedding, β_t is used to individually predict each word, k, in the window using a softmax classification between β_t and $\beta_k \forall k \in w_\Delta$. The parameters for the embeddings are then fitted by minimizing the cross-entropy loss using stochastic gradient descent. A graphical depiction of this process is in Figure 2.

Figure 2: Model Architecture with Window of 6: Word Embedding Estimation



The second part of the model trains a document vector, ζ_i, for each candidate-year, i. This model architecture is the same as the first sept, but instead of using a word embedding to predict words, the candidate embedding, $zeta_i$, replaces the word embedding for the target

Figure 3: Model Architecture with Window of 6: Candidate-Year Embedding Estimation



word and is used to predict the words in w_{Δ}. Intuitively, this means candidate embeddings are trained to have parameter weights that reflect the word embeddings in candidates' issue statements. Like the first step, the parameters for the embeddings are fitted by minimizing the cross-entropy loss using stochastic gradient descent. A graphical depiction of this step is provided in Figure 3. It should be noted that while various model architectures exist, the one used in the body of the paper follows best practices for Doc2Vec implementations (for example, see Lau and Baldwin 2016).

Model Implementation

Before fitting the word embedding model on campaign website text, it is important to discuss a number of parameter-level decisions in creating the resulting WEB Scores. Starting with text pre-processing, I follow the same procedure as Rodriguez and Spirling (2022) and convert all tokens to lower case and remove all non-text characters. In addition, I also remove words that do not appear across the full set of documents more than five times. This is done because Doc2Vec uses an estimation strategy that generally over-weights rare terms in the training process.[22] Removing infrequent terms improves the accuracy and performance of the models (Rodriguez and Spirling 2022).

I fit the model using parameter recommendations from Rodriguez and Spirling (2022), including a window of 6 and an embedding dimension of 300.[23] In addition, I also use pre-trained Word2Vec embeddings from the Google News corpus. This is done due to the limited data from campaign issue statements for training a word embedding model and ensures high-quality word embeddings are used in the training process. The use of pre-trained embeddings also improves the performance of Doc2Vec embedding models overall Lau and Baldwin (2016). Finally, I use default hyperparameter recommendations from Mikolov, Yih and Zweig (2013) with an increased number of epochs (20). The larger number of epochs is

[22]Estimates of candidate positioning are highly correlated with different cutoff thresholds (0, 5, 10, 20). See Table 1 for correlations of different hyper-parameter specifications

[23]Because there is no clear-cut justification for model parameters, I also fit models with various window sizes (5, 6, 7, and 8) and embedding dimensions (100, 200, 300) and show that resulting measures are almost perfectly correlated (≥ 0.99), suggesting parameter decisions have little effect on the resulting scores. See Table 1 for correlation tables from different model architectures.

due to a limited number of documents per candidate-year occurrence and is consistent with Rheault and Cochrane (2020).

Model Robustness

Table 1: Candidate Positioning Correlation Table with Different Model Parameters (window size, embedding dimension)

	5, 100	5, 200	5, 300	6, 100	6, 200	6, 300	7, 100	7, 200	7, 300	8, 100	8, 200
5, 200	0.995										
5, 300	0.994	0.996									
6, 100	0.997	0.995	0.994								
6, 200	0.995	0.997	0.996	0.995							
6, 300	0.993	0.996	0.997	0.993	0.996						
7, 100	0.997	0.994	0.993	0.997	0.994	0.993					
7, 200	0.994	0.997	0.996	0.995	0.997	0.996	0.994				
7, 300	0.993	0.996	0.996	0.993	0.996	0.997	0.993	0.996			
8, 100	0.996	0.994	0.992	0.997	0.994	0.993	0.997	0.994	0.993		
8, 200	0.994	0.996	0.995	0.995	0.997	0.996	0.995	0.997	0.996	0.994	
8, 300	0.992	0.995	0.996	0.993	0.996	0.997	0.993	0.996	0.997	0.993	0.996

C Full Issue Statements for Internal Validity Test

Table 2: Full Issue Statements from Heritage Foundations

Policy Area	**Issue Statement**
Abortion	All children conceived deserve to be born to married mothers and fathers who will love, guide, and protect them throughout their lives, but **family** breakdown and rampant abortion have torn apart the soul of our country and sapped it of its strength and moral authority. We will advance the **Heartbeat** Protection Act to **prohibit** abortion nationwide after the moment a heartbeat can be detected. At the state level, we will work with governors, legislators, and other state-based allies to pass heartbeat laws (or better) on abortion. We will work to **prohibit** the interstate commerce of abortion pills in **pro-life** states by advancing legislation in both the House and Senate.
Education	The Heritage enterprise will work to minimize the federal government's intervention in education. The education system is failing our children—from the scourge of woke ideas like critical race theory and radical gender ideology to the lack of accountability to **parents** and an absence of academic **transparency**. **Parents**, not bureaucrats, should be making teaching and learning decisions that align with their values. Taxpayer dollars should help students to succeed with a great education, not prop up failing school systems. The entire Heritage enterprise will spearhead reforms at the state level to protect parental rights and expand education **choice** and will work at the federal level to limit Washington's intervention, ultimately driving a clarion call to eliminate the U.S. Department of Education. Minimizing federal intervention in education includes supporting the introduction of federal legislation to (1) give states more budget authority over federal education funding with fewer strings, (2) reduce federal intervention in early childhood education by reforming programs such as Head Start, and (3) expand families' access to homeschooling by reforming 529 savings accounts to include **homeschooling** expenses and by expanding and making permanent the D.C. Opportunity Scholarship Program.
Government	Government **spending**, **regulations**, and inflation are a tax on all Americans, especially working families who struggle to make ends meet. **Prudent fiscal** decisions by government can enable American families to flourish without politicians and bureaucrats controlling their lives. The Heritage enterprise will advance a blueprint to reduce the size and scope of the federal government, ensure that government spends less of our money to save us from falling off the **fiscal** cliff, and stop the growth of federal **regulations**.
Immigration	Americans should be able to live peacefully without constant fear of crime or **incursions** across our borders. A strong justice system **enforces** existing U.S. law, **prosecutes** criminals, **secures** our borders, and preserves our national identity. America's current border crisis and the level of crime in many cities are out of control, and the human costs are staggering.

Note: Keywords from Table 3 are bolded in each issue statement.

Table 3: Full Issue Statements from Justice Democrats

Policy Area	**Issue Statement**
Environment	Now is the time for a comprehensive, once-in-a-generation mobilization that prioritizes front-line communities, combats the **climate** crisis, and creates millions of good-paying union jobs. A Green New Deal will transition away from **fossil** fuels and dramatically expand existing **renewable** power sources with the goal of meeting 100% of national power demand through **renewable** sources. A Green New Deal also provides people across the country with the opportunity, training and education needed to participate fully and equally in a green economy, offering jobs to help rebuild our crumbling infrastructure. A Green New Deal ensures a just transition for all workers, low-income communities, communities of **color**, and indigenous communities.
Guns	Gun violence is a public health crisis in the United States that disproportionately impacts communities of color. More than 90 percent of Americans support expanded **background** checks, 54 percent want a **ban** on **assault** weapons and 54 percent want a ban on high capacity magazines. We agree with the majority of the American people and support these measures. To enact common sense gun **safety** measures, we must break the NRA's hold on our corrupt government and prioritize the mental and physical health of the people over the billion-dollar gun manufacturing industry's bottom line.
Healthcare	The United States has the most expensive and least effective healthcare system compared with other industrialized nations. It's time to end the destruction of healthcare in America by price gouging, for-profit private health insurers and catch up to every other modern nation that's implemented a **single-payer universal** healthcare system – no networks, no premiums, no co-pays, no deductibles and no surprise bills. **Medicare-For-All** will **expand** Medicare coverage to include dental, hearing, mental health and substance abuse treatment, prescription drugs, long-term and disability care, and reproductive and maternity care. We must also invest in frontline care workers who are the backbone of our economy.
Wages	Over the past several decades, the **cost** of **living** has increased significantly while workers' wages have remained relatively stagnant. While CEO's compensation soars, most workers' wages aren't even keeping up with inflation and **affordable** housing remains out of reach. We must secure a **minimum** wage of at least $15 that's tied to inflation.

Note: Keywords from Table 3 are bolded in each issue statement.

D Internal Validity Test Regression Tables

Table 4: Full Regression Tables from Figure 4 for Conservative Policies

	Dependent variable:			
	Government	Immigration	Abortion	Education
	(1)	(2)	(3)	(4)
WEB Scores	0.026***	0.016***	0.006***	0.011***
	(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.261***	0.326***	0.366***	0.310***
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	4,554	4,554	4,554	4,554
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 5: Full Regression Tables from Figure 4 for Liberal Policies

	Dependent variable:			
	Environment	Healthcare	Guns	Wages
	(1)	(2)	(3)	(4)
WEB Scores	−0.027***	−0.031***	−0.014***	−0.024***
	(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.325***	0.262***	0.399***	0.512***
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	4,554	4,554	4,554	4,554
Note:				*p<0.1; **p<0.05; ***p<0.01

E Table 4 Robustness Checks

Table 6: Replication of Table 4 Column 1 Using Cross-Pressure Measure

	Measurement
	WEB Scores
Moderate Challenger	-0.062
	(0.040)
Extreme Challenger	0.099^{***}
	(0.036)
Constant	0.011
	(0.026)
Observations	518
Candidate Random Effects	✓
Year Fixed Effects	✓
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 7: Replication of Table 4 Column 1 with Flipped Coding for Cross-Pressured Incumbents

	Measurement
	WEB Scores
Extreme Challenger	0.067^{*}
ref: No Challenger	(0.039)
Moderate Challenger	-0.052
ref: No Challenger	(0.041)
Constant	0.016
	(0.027)
Observations	518
Candidate Random Effects	✓
Year Fixed Effects	✓
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 8: Replication of Table 4 with Fixed Effects

	Measurement:			
	WEB Scores	CFscores	WEB Scores	CFscores
	(1)	(2)	(3)	(4)
Extreme Challenger	0.074	0.028		
	(0.050)	(0.021)		
Moderate Challenger	-0.117^{**}	0.066		
	(0.044)	(0.043)		
Challenger Extremity			0.142	0.031
			(0.110)	(0.044)
Constant	-0.060	-0.178^{**}	0.250	-0.064
	(0.162)	(0.089)	(0.235)	(0.091)
Observations	518	466	189	122
Candidate Fixed Effects	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 9: Replication of Table 4 with 2018, 2020, and 2022 using WEB Scores

	Measurement:			
	(1)	(2)	(3)	(4)
Extreme Challenger			0.141^{***}	0.083
			(0.046)	(0.071)
Moderate Challenger	-0.033	-0.009		
	(0.033)	(0.039)		
Challenger Extremity			0.142	0.031
			(0.110)	(0.044)
Constant	0.023	0.378	0.048	0.369
	(0.025)	(0.241)	(0.039)	(0.244)
Observations	795	795	315	315
Candidate Random Effects	✓		✓	
Candidate Fixed Effects		✓		✓
Year Fixed Effects	✓	✓	✓	✓
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table 10: Replication of Table 4 with Consistent Sample

	Measurement:			
	WEB Scores	CFscores	WEB Scores	CFscores
	(1)	(2)	(3)	(4)
Extreme Challenger	0.006	−0.026		
	(0.042)	(0.022)		
Moderate Challenger	−0.052	0.033		
	(0.065)	(0.044)		
Challenger Extremity			0.192**	−0.006
			(0.093)	(0.038)
Constant	0.023	−0.273***	−0.004	−0.345***
	(0.027)	(0.020)	(0.061)	(0.045)
Observations	428	428	102	102
Candidate Fixed Effects	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓
Note:				*p<0.1; **p<0.05; ***p<0.01