Measuring Strategic Positioning in Congressional Elections

Short Title: Strategic Positioning in Congressional Elections

Colin R. Case*

August 30, 2025

Abstract

How does an incumbent's issue positioning respond to an extreme (moderate) primary challenger? While theoretical models of electoral competition suggest incumbents should adopt more extreme (moderate) positions, prior empirical work testing this hypothesis does not find support for this hypothesis. I argue existing measures of campaign positioning are not suited to adequately test this hypothesis. To overcome these data limitations, I introduce Website EmBedding (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.

Keywords: Polarization, Congress, campaigns, text analysis

*email: colin-case@uiowa.edu. Replication files are available in the *JOP* Data Archive on Dataverse (`https://dataverse.harvard.edu/dataverse/jop`). The empirical analysis has been successfully replicated by the *JOP* replication analyst. Supplementary material for this article is available in the appendix in the online edition.

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 positions to the left (e.g., "bolder action to protect our environment"); Kim won the primary decisively, with almost 93% of the vote. Prior work has placed a substantial focus on understanding how incumbent candidates change their behavior in response to primary challengers, such as changes in ideological self-placement (Burden 2004), party unity voting in Congress (Jewitt and Treul 2019), domains of news stories shared on social media (Macdonald et al. 2025), and partisanship of language on social media (Cowburn and Sältzer 2025), among others. In addition, theoretical accounts predict that incumbent candidates should change their overall issue positioning in response to a primary challenger. However, prior work finds no association when it comes to a change in aggregate incumbent positioning (e.g., Boatright 2014; Hirano and Snyder 2019).

Why has prior research found a lack of support for incumbents responding to the positioning of primary challengers? I argue that measurement limitations have prevented scholars from fully testing how incumbents position themselves in response to primary challengers. The aforementioned research places a substantive focus on roll-call behavior and does not consider the issues candidates run on in response to an extreme primary challenger. While roll-call voting is an important component of members' behavior in Congress, understanding members' issue positioning on the campaign trail has implications for questions related to voter decision making (e.g., Hassell and Visalvanich 2024), democratic accountability and representation (e.g., Bafumi and Herron 2010; Bonica and Cox 2017), polarization (e.g., Boatright 2014; Hirano and Snyder 2019), and future legislative (e.g., Sulkin 2005; Grimmer 2013). The effect of primary challengers on incumbent positioning, and subsequently these related questions, carries implications for a large proportion of congressional races. In each election cycle, races where incumbents face a primary challenger represent a sizable proportion of all congressional races (21% of all partisan primaries in 2018 and 2020; 40% of all congressional districts in states with partisan primaries in 2018 and 2020). In this way, understanding how primary election challengers shape incumbent positioning has broad implications for understanding a range of important questions in American politics.

Despite its importance, existing measures of campaign positioning are not suited to test incumbents' positioning response to primary challengers for two reasons. First, existing measures of campaign positioning are often approximations based on either related candidate behaviors or citizen perceptions' of candidates' positions. In the case of the first type of measurement, it is not clear the durability or the broader implications of these changes. In the case of the second type of measurement, changes in scores using approximations could be partly due to groups of citizens', such as campaign contributors, responding to changes in electoral dynamics rather than changes in candidate behavior. Simply put, existing measures of campaign positioning are not based on the underlying data researchers are interested in quantifying: the issues candidates run on during the campaign. Second, 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; Hare et al. 2015; Barberá 2015; Macdonald et al. 2025; Gaynor et al. 2025) 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. 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, Case and Treul (2025) 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). In addition, campaign website issue positions do not change from the primary to the general election (Porter, Treul and McDonald 2024). In this way, campaign website issue positions capture the trade-off candidates must make in their issue positioning to balance electoral considerations between the primary and general election. Finally, they also cover a substantively important concept of interest related to candidate positioning – the issues candidates *actually* run on.[1] In this manner, WEB Scores are conceptually distinct from other positioning measures (e.g., CFScores) in that they explicitly measure strategic candidate positioning, rather than approximations.

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 all primary candidates with campaign website issue positions, I show the resulting measure has high validity and improves on widely used measures of candidate positioning. First, it greatly expands the coverage of congressional candidates (75.1% versus 64.3%). Second, it better captures the actual quantity of interest: the overall positioning of candidates' policy proposals during the campaign. Using this measure, I show incumbent candidates challenged by an extreme (moderate) challenger become more extreme (moderate) in their positioning during campaigns. Consistent with prior work and my theoretical expectations, I do not find the same

[1] Campaign websites are just one component of candidates' communication during the election. Candidates can communicate with voters through social media, advertisements, speeches, and more. While candidates can offer issue positions through these other formats, they are incentivized to run on consistent positions across their campaigns. In this manner, campaign websites offer an unmediated format for candidates' issue positions, likely reflected in these other formats.

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 positioning. Although campaign positions are not directly tied to legislating, this campaign behavior still has implications for future legislative action (Sulkin 2009; Grimmer 2013; Sulkin 2005). Second, I provide an off-the-shelf measure of candidate positioning for researchers that (1) better captures the actual issues candidates run on during the campaign and (2) increases the number of candidates covered. 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.

Campaign Issue Positioning

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 running for office (Thomsen 2014). One question of particular attention has looked at whether or not incumbents respond to extreme primary challengers by changing their overall issue positioning.

In two-party competition along a single-issue dimension and a one-stage election, theoretical accounts predict that electorally motivated candidates will converge to the median voter (Downs 1957). In congressional elections with a partisan primary, candidates face competition at two stages. Candidates first compete against co-partisan candidates and appeal to a smaller, more partisan subset of the electorate. The winner moves to the general election, competes against out-partisan candidates, and must appeal to a broader electorate.

It is important to note candidates are largely constrained to offering the same issue positioning for both the primary election and the general election (Coleman 1971; Aranson and Ordeshook 1972; Owen and Grofman 2006; Cowburn and Sältzer 2025); changing issue positions within an election cycle could result in an electoral penalty due to flip-flopping (Canes-Wrone, Brady and Cogan 2002; Gooch 2022). Incumbents, therefore, need to adopt issue positions that balance the strategic considerations between both the primary and the general election.

Unlike models of single-stage elections, formal models of two-stage elections predict that candidates adopt positions that do not converge to the median voter. As Coleman (1971) notes, candidates running in a primary election must first win over non-centrist party voters. Given this, formal models with various assumptions predict candidates' positions are pulled towards the party median rather than the median voter because they must appeal to partisan voters first during primary elections (Coleman 1971; Aranson and Ordeshook 1972; Owen and Grofman 2006). One implication of these models is that the positioning of the primary challenger matters for the strategic positioning considerations of the incumbent. If a moderate candidate challenges an incumbent, incumbents can shift their positions towards the median voter in the general election while still retaining a majority share of the primary electorate. If an extreme candidate challenges an incumbent, then formal models predict candidates adopt more extreme positions to appeal to non-centrist party voters (Owen and Grofman 2006).[2]

[2]The aforementioned theory assumes there is a single primary challenger. However, the above theory and formal models can be applied to circumstances with more than one primary challenge. If all primary challengers are either moderate or extreme, incumbents can respond in the same direction to these challengers. In the case where one or more challengers are moderate, and one or more challengers are extreme, the expectation is less clear. It may be the case that the incumbent responds to the most threatening candidate. It could also be the case that the incumbent balances the challenge from both. In the main analysis of the paper, I assume that incumbents will respond to the most threatening challenger. However, I show the results of this paper are robust for testing incumbents who are cross-pressured (challenged by both moderate and extreme candidates). It should be noted that this situation only happens in a small percentage of primary elections: 21% of incumbents face multiple primary challengers, and only 7% face cross-pressured primary

While not testing these formal models explicitly, prior empirical work does demonstrate incumbent candidates are responsive to primary election competition. For example, Burden (2004) demonstrates candidates' ideological self-placement on mail surveys is more ideologically extreme when facing a primary challenger. Other work by Macdonald et al. (2025) demonstrates the domain source of news stories incumbents share on Twitter changes in response to a primary challenger. Cowburn and Sältzer (2025) also finds that incumbents' Twitter content becomes less partisan after losing to a primary challenger. Given both the theoretical expectations and empirical evidence that incumbent candidates are responsive to primary election competition in other aspects of their campaign, I argue the following:

Hypothesis: Incumbent candidates adopt more extreme (moderate) issue positions in response to an extreme (moderate) primary challenger.

A number of scholars have tested this hypothesis previously. For example, Boatright (2014) looks at different types of primary challenges (e.g., ideologically extreme 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. While prior work suggests incumbent candidates do not respond to the positioning of primary challengers, I argue existing work uses poorly-suited measures to test this hypothesis.

One of the commonalities across research assessing whether or not incumbents' positioning changes in response to extreme challengers is the focus on measures of positioning based 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. While a useful measure of *legislative* behavior, measures such as NOMINATE are not suited to capture changes in incumbent *campaign* positioning: members only vote on the select issues that make it to the floor, not the universe of all issues, and scores are often influenced votes on procedural matters (Roberts 2007) and votes with no issue content (Lee

challenges.

2009). In addition, in an era of strong political parties, few issues come to the floor that divide parties internally, masking potential intra-party conflict (Cox and McCubbins 2005).

Instead, I argue changes in incumbent positioning should be observed in the issue positions candidates take during the campaign. Unlike roll-call voting, campaign issue positions are not constrained by a legislative agenda; incumbents can take nuanced positions to differentiate themselves from co-partisan candidates. Moreover, these campaign issue positions have important implications for the broader context of American politics. Politicians tend to make good on their campaign promises (Ringquist and Dasse 2004; Sulkin 2009, 2011; Meinke 2023) and translate these campaign issue positions to future legislative priorities (Sulkin 2005; Grimmer 2013). Campaign behavior can also signal a politician's future legislative style (Fenno 1978; Sulkin and Swigger 2008). In addition, incumbent campaign position matters when it comes voter decision making (Hassell and Visalvanich 2024, e.g.,), democratic accountability and representation (e.g., Bafumi and Herron 2010; Bonica and Cox 2017), and polarization (e.g., Boatright 2014; Hirano and Snyder 2019). Despite the broader importance of campaign positioning, existing measures are not well suited to test incumbents' responses to primary challengers.

Generally speaking, measures of campaign positioning fall into one of two typologies: citizen perceptions of candidates or actual candidate behavior. Within the first typology, measures based on citizen perceptions assume that citizens can consider various aspects of candidate positioning, such as the issues they run on, policy goals, and values (Bonica 2014). Common approaches often ask survey respondents (Hare et al. 2015; Ramey 2016) or experts (Hirano et al. 2015) to place candidates spatially from liberal to conservative and then aggregate these responses to position candidates using various scaling methods. Other approaches rely on aggregating a subset of citizens' behaviors, such as donors (Bonica 2014) or followers on Twitter (Barberá 2015); these estimation strategies assume that citizens donate to and follow candidates on Twitter positioned proximal to one another.

The second typology of measurement approaches approximates candidate positioning using other related candidate behaviors. For example, Macdonald et al. (2025) 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. (2025) and Cowburn and Sältzer (2025), 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 in studying incumbents' positioning 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 congressional floor speeches, are limited to incumbent candidates.[3]

[3] It should be noted that approaches using candidates' Tweets could get around this limitation, but existing research has only collected data on subsets of candidates (e.g., Gaynor et al. 2025) or for a single election year (e.g., Cowburn and Sältzer 2025). Additionally, Twitter and other social media data do not fully reflect the single-position constraint between the primary and the general election. Unlike a website issue page where candidates' positions are presented simultaneously, social media data is temporal across the election cycle. While there is value in this type of temporal data (for example, see Macdonald et al. 2025; Cowburn and Sältzer 2025), it does not provide a stable picture of aggregate candidate positioning across the full election cycle. Instead, candidates can emphasize certain issues that cater to the electorate (primary versus general) they are appealing to without flip-flopping their issue positions. Further, employing social media data has also become more complicated and costly with recent roll-backs to academic researcher access.

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 many 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 2025). It should be noted that while very few incumbent candidates actually lose to these noncompetitive primary challengers, incumbents are still wary of these challenges.[4] As such, measuring these challengers' issue positioning is a crucial component of understanding incumbent behavior. For context, CFscores, which provide some of the highest levels of coverage of primary candidates among existing measures, does not have a score for 64.3% 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 or Tweets, fail to capture the actual issue positions candidates take during the campaign. This is problematic for adequately testing incumbents' changes in issue positioning in response to a primary challenger. In the case of the first measurement typology that relies on citizen perceptions, it could be the case that changes in measurement are due to citizens' response to changing electoral circumstances, not changes in incumbent behavior. 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.

[4] For example, in 2013, former Speaker of the House Dennis Hastert remarked "It used to be [members are] looking over their shoulders to see who their general [election] opponent is. Now they're looking over their [shoulders] to see who their primary opponent is... And so everybody's kind of neurotic about where their support is."

In the case of the second measurement typology that focuses on related candidate behaviors, the broader implications of changes in behavior (e.g., what links are shared on Twitter) are not well understood. It could be the case that existing measures, such as those based on the content of social media posts, captures short-term changes in behavior rather than enduring changes in issue positions that carry over to future legislative behavior.

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). As evidence of this, over a dozen states include links to campaign websites provided by candidates on official listings of ballot-eligible candidates for voters, donors, and journalists to access in a centralized hub. 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 (Porter, Treul and McDonald 2024), in large part to avoid flip-flopping on issue positions. As a result, when candidates put their websites together, they must balance considerations between the primary and the general election to avoid changing the content of their issue positions between the elections. Campaign websites, therefore, reflect the positional constraint placed on candidates across an election cycle and are a comprehensive data source for studying candidates in U.S. congressional elections.[5]

[5]It could be the case that campaign websites are constructed with the thought of warding off a potential primary challenger. However, candidates challenging incumbents register early in the election cycle, making

As part of their campaign website, most (75%) 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, Case and Treul (2025) 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. 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.[6] 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 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 a candidate's primary election date to ensure consistency in the data collection process and that candidates' websites were finalized in the lead-up to the election. This data set contains 4,509 issue pages (75.1% of all candidates; 85% of candidates with a website).[7]

it unlikely campaign websites are used by incumbents to ward off primary challengers. To demonstrate this, I look at when candidates challenge incumbents in the primary and register with the FEC, a common indicator of the start of a candidates' campaign that does not require fundraising (Bonica 2020). On average, these candidates with the FEC are more than 8 months (240 days) before the primary election. This places the timing of these registrations in September of the year prior to the election, giving very little time for incumbents to try and prevent primary challengers from their last election. It is more likely that candidates are aware of these primary challenges when crafting their website for the primary election, given the timing.

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

[7] 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 operating their campaigns, 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

Campaign issue pages improve upon current measurement approaches through both the expansion of the number of candidates included and by 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 2018-2022 U.S. House of Representatives primaries. This is further broken down by candidate type: incumbents, non-incumbents who have previously held elected office, and non-incumbents who have not previously held elected office. In the aggregate, 4,509 (75.1%) candidates had an issue page on their campaign website in 2018-2022 and 3,864 (64.3%) have a CFscore.[8] As is evident in Figure 1, campaign websites provide a large increase in coverage of candidates when it comes to inexperienced candidates. Of the 4,085 inexperienced candidates who ran in 2018-2022, 2,882 (70.5%) had an issue page on their campaign website, while only 2,087 (51.1%) received enough eligible contributions for a CFscore. When it comes to experienced challengers, both sets of data have a high percentage of candidates, with 593 (77.0%) having an issue page and 630 (81.1%) out of 770 total experienced challengers having a CFscore. 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 (90%) than CFscores (100%).

Estimation Strategy

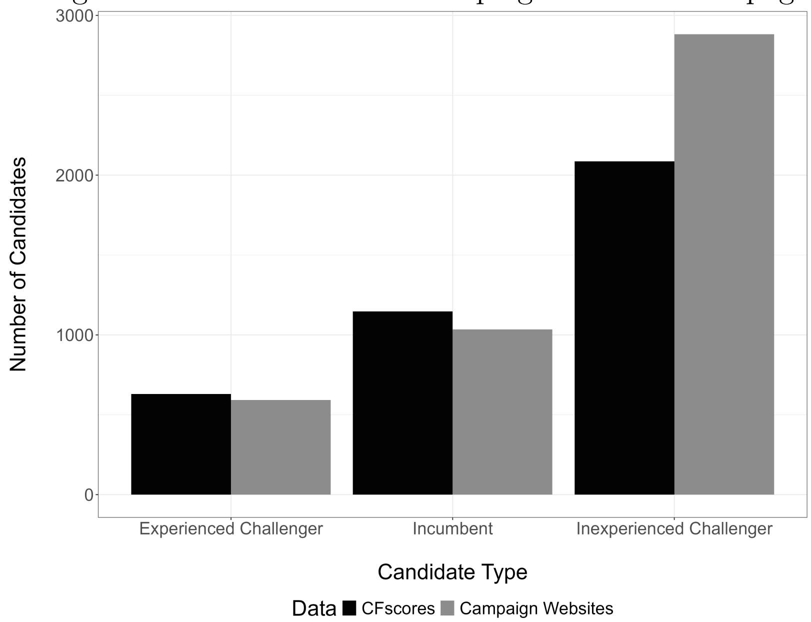
To estimate Website EmBedding (WEB) Strategic Positioning Scores using issue position text, I rely on a word embedding model with document-level vectors (Doc2Vec; for the original model specification, see Le and Mikolov 2014) for each candidate-year. Word embeddings

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.

[8] 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.

Figure 1: Candidate Coverage by Measurement and Candidate Type

Alt Text: Bar plot showing the number of experienced challengers, incumbents, and inexperiend challengers with a CFScore and campaign website issue page, respectively.



Note: Figure 1 depicts the number of candidates running as either a Democrat or Republican in 2018-2022 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 without previous elected experience.

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 widespread 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 2023) 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 can 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.

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 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. In many ways, this process is similar to WordFish (Slapin and Proksch 2008). However, unlike previous approaches, word embeddings capture meaning (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.

A common challenge of estimating positioning from text is the extent to which the resulting scores capture issue positions (e.g., how extreme or moderate is a candidate's position) and issue emphasis (e.g., the quantity of extreme or moderate positions a candidate takes), or both. There is a conceptual argument for how both components relate to more extreme scores. The estimation procedure described above incorporates both components into the resulting estimation. In Appendix C, I use altered candidate websites to demonstrate the relationship between WEB Scores and issue emphasis. It is important to note that this does not mean more issue text results in more extreme WEB Scores; the correlation between WEB Score extremity and the amount of text from each candidate is -0.06.

Measurement Validity

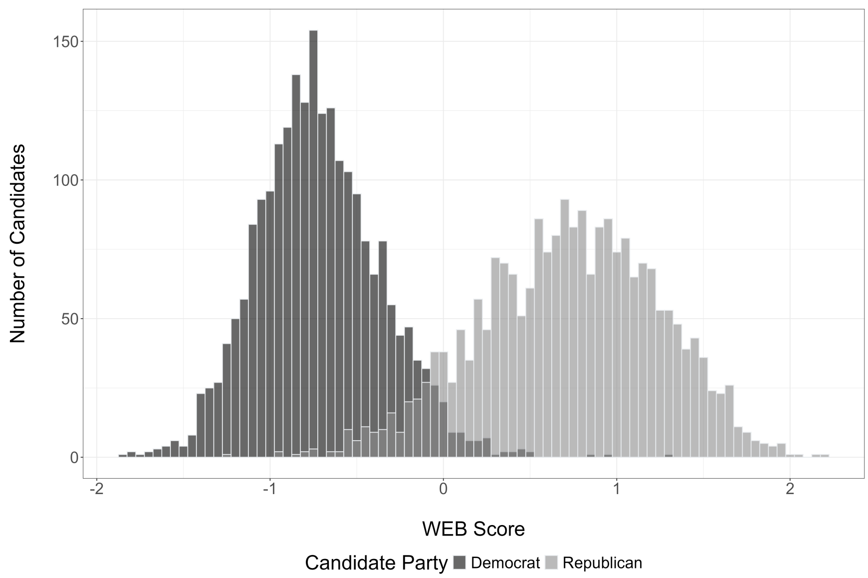
In this section, I provide an overview of the measurement and several validation procedures. The distribution of WEB Scores is plotted in Figure 2. The measurement has a mean of 0 across candidates and a standard deviation of 0.85. Democratic candidates trend to the negative side of the scale with a mean score of -0.72 while Republicans have a mean score of 0.72. Unlike other measures of campaign positioning (e.g., DW-NOMINATE, CFScores), there is significant overlap between candidates from the two major political parties. This overlap is likely due to the reduced influence of partisanship on the data-generating process; while parties do influence campaign issues, candidates have more degrees of freedom in the issue positions they take.

To provide face validity 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 (2022), sit well to the left of the Democratic mean, with a score of -1.10. On the Republican side, Marjorie Taylor Greene (2020) also has a score well to the extreme of the Republican mean at 1.68 in her first year running for office.

Next, I turn to evaluating the similarity between WEB Scores and pre-existing scores of positioning: CFscores, which scale based on donors' perceptions of candidates' positioning, and DW-NOMINATE, which scale based on voting preferences on the congressional legislative agenda. It should be noted that while these concepts are distinct from explicit

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

Alt Text: Histogram showing the distribution of candidates' WEB Scores broken down by party.



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.

candidate positioning, they should nonetheless be somewhat related albeit not perfectly correlated (Bonica 2014). Table 2 shows the correlations incumbent candidates running in 2018-2022, restricted to those having a WEB Score, a CFscore, and a DW-Nominate score.[9] 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 0.90, as well as the correlation between WEB Scores and CFscores at 0.88. These correlations are substantively similar to the correlation between DW-Nominate and CFscores at 0.95. The high correlation between all three measures is largely a function of the measures separating the two political parties.

Turning to intra-party correlations for Democratic candidates, WEB Scores are weakly correlated with DW-Nominate at 0.23 and with CFscores at 0.29. Both are significantly

[9]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 0.87, Democratic candidates is 0.20, and Republican candidates is 0.21.

Table 1: Most Liberal and Conservative Incumbent Candidates (2018-2022)

Most Liberal	Most Conservative
Katherine Clark (2022, -1.48)	Troy Bladerson (2020, 2.18)
Brenda Lawrence (2018, -1.45)	William Timmons (2020, 1.97)
Ilhan Omar (2022; -1.42)	William Timmons (2022, 1.96)
Pramila Jayapal (2020, -1.42)	Jeff Duncan (2022, 1.93)
Ilhan Omar (2020, -1.41)	Neal Dunn (2018, 1.89)
Mondaire Jones (2022, -1.34)	Neal Dunn (2022, 1.85)
Pramila Jayapal (2022, -1.33)	Jeff Duncan (2020, 1.82)
Melanie Stansbury (2022, -1.32)	Debbie Lesko (2020, 1.77)
Yvette Clark (2022, -1.30)	Troy Balderson (2018, 1.77)
Pramila Jayapal (2018, -1.29)	Warren Davidson (2022, 1.75)

Note: Table shows the most liberal and conservative incumbent candidates who ran in the 2018, 2020, and 2022 congressional elections. Incumbents' WEB Score is included in parentheses after the election year.

higher than the correlation between CFscores and DW-NOMINATE for Democrats (0.08). Among Republican candidates, the correlation between WEB Scores and DW-NOMINATE is moderate at 0.42. This is lower than the intra-party correlations for CFscores and DW-Nominate at 0.60. The correlation between WEB Scores and CFscores is weak at 0.29.

In the aggregate, these correlations provide evidence the measures are capturing related but distinct concepts, as expected. Given the empirical distinction, it is important to consider how these various measures capture the quantity of interest: the positional leaning of issue positions from the underlying text. Human judgments of political text represent the "gold standard" for validating measures of positioning created from text (Grimmer and Stewart 2013). For this reason, I compare WEB Scores with human judgments of candidates' issue statements to validate WEB Scores as a measure of campaign positioning. I also compare the performance of WEB Scores with other measures of positioning (CFScores and DW-NOMINATE) and show WEB Scores better capture human judgments of issue position text.

Hand-labeling large amounts of text is both costly and time-intensive. Large language models are well suited to accomplish human labeling tasks and produce similar results to human coders, including labeling the ideological scaling of political texts (Le Mens and Gallego 2025; Ornstein, Blasingame and Truscott 2025). The primary benefit of using GPT-4 for labeling texts is both cost and time efficiency. This allows me to label every issue

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

All Members of Congress			
	CFscores	DW-NOMINATE	WEB Scores
CFscores	1.00	–	–
DW-NOMINATE	0.95	1.00	
WEB Scores	0.89	0.89	1.00
Democrats			
	CFscores	DW-NOMINATE	WEB Scores
CFscores	1.00	–	–
DW-NOMINATE	0.08	1.00	–
WEB Scores	0.29	0.23	1.00
Republicans			
	CFscores	DW-NOMINATE	WEB Scores
CFscores	1.00	–	–
DW-NOMINATE	0.60	1.00	–
WEB Scores	0.29	0.42	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.

statement from candidates instead of just a small subset.

To label statements, I used the original issue statement text from Porter, Case and Treul (2025). In doing so, I used the R package promptr's "complete chat" function, which allows users to interface with OpenAI's API through R. These chat models are better suited to labeling text with zero training examples (Ornstein, Blasingame and Truscott 2025). To label each individual text, I used an adapted prompt for ideological positioning as those used in Le Mens and Gallego (2025) and Ornstein, Blasingame and Truscott (2025). The instructions for GPT read as follows: "You will be provided with a text from a candidate running for the U.S. Congress. Where does this text stand on the 'liberal' to 'conservative' scale? Provide your response as a score between 0 and 100, where 0 means 'Extremely liberal' and 100 means 'Extremely conservative'. Respond only with your score." For each issue statement, the instructions were sent along with the individual issue statement. GPT-4 would then return up to ten possible tokens for the ideological positioning, each an integer between 0 and 100. I then assigned the token with the greatest predicted probability as the

Table 3: Correlations with GPT-4 Labeled Issue Statements

Candidates with a CFScore and WEB Score			
	All Candidates	Democrats	Republicans
WEB Scores	0.93	0.61	0.70
CFScores	0.90	0.27	0.24

Candidates with a DW-NOMINATE Score and WEB Score			
	All Candidates	Democrats	Republicans
WEB Scores	0.94	0.62	0.73
DW-NOMINATE	0.92	0.35	0.41

Note: Table 3 shows the correlation coefficient between GPT-4 generated scores ("human" labels) and existing measures of positioning. The top panel only includes candidates with a CFScore and a WEB Score. The bottom panel only includes candidates with a DW-NOMINATE score and a WEB Score.

score for the issue text.[10]

To aggregate scores to the candidate level, I then averaged all scores across each individual issue statement for candidates. This produced a single score for each candidate-year observation from 0 to 100.[11] I then generated correlations between GPT-4 generated scores and existing measures of campaign positioning. Table 3 presents correlations with GPT-4 generated scores for WEB Scores and CFScores in the top panel and correlations with GPT-4 generated scores for WEB Scores and DW-NOMINATE in the bottom panel. Correlations for all candidates are in the left column, Democratic candidates in the middle column, and Republican candidates in the right column. To ensure I can make direct comparisons between measures, I only include candidates with both a CFScore and a WEB Score in the top panel and only candidates with DW-NOMINATE and a WEB Score in the bottom panel. Correlations for all candidates with an issue page are consistent with those in Table 3 (0.92 for all candidates, 0.63 for Democratic candidates, and 0.72 for Republican candidates).

Starting with all candidates, correlations are high between GPT-generated labels and

[10]To validate that GPT-4 reflects expert human coding, I labeled a random sample of 200 statements using the same instructions. My labeling and GPT's labeling were highly correlated (> 0.75 within party correlations), further validating the use of GPT-4 for human labeling tasks as shown in Ornstein, Blasingame and Truscott (2025) and Le Mens and Gallego (2025). Full coding instructions and correlation tables for the hand labeling can be found in Appendix D.

[11]Results are consistent if I also produce a weighted average score by the amount of text in the issue statement.

measures of positioning. However, these high correlations are primarily a function of party differences in scores; when validating positioning measures, it is necessary to focus on intra-party correlations (Tausanovitch and Warshaw 2017). WEB Scores capture GPT-generated scores significantly better than both CFScores (0.61 versus 0.27) and DW-NOMINATE (0.62 versus 0.35) for just Democratic candidates. The same trend occurs with Republican candidates, with WEB Scores having significantly higher correlations than CFScores (0.70 versus 0.24) and DW-NOMINATE (0.73 versus 0.41).

In totality, the high correlations between WEB Scores and GPT-generated labels, both across and within parties, demonstrate (1) WEB Scores well-capture campaign positioning in candidate issue statements and (2) better capture campaign positioning in candidate issue statements than alternative measures of positioning, as measured by GPT-generated human judgements. In many ways, this result should not be surprising; WEB Scores are generated from the exact text that GPT-generated labels are based on. But to the extent that GPT-generated labels reflect human perceptions of issue text, the results further show existing measures of positioning (CFScores, DW-NOMINATE, and WEB Scores) are distinct quantities of interest. In the appendix, I carry out supplemental validation tests. In Appendix E, I demonstrate the external validity of WEB Scores and show they capture differences in congressional ideological caucuses. In Appendix F and G, I also show that WEB Scores capture word relationships between word embeddings, demonstrating measure is picking up on various policy proposals associated with the liberal and conservative endpoints in text.

Analysis

Given the validity of WEB Scores and 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 using an incumbent candidate's WEB Score, where I 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 adopting more extreme (moderate) issue positioning. Importantly, this dependent variable captures candidates' overall positioning for an election cycle. As mentioned previously, candidates face an electoral penalty for flip-flopping within an election cycle (Canes-Wrone, Brady and Cogan 2002; Gooch 2022). The content on their issue page and, subsequently, their WEB Scores are constrained within an election cycle because of this. As a result, the dependent variable captures incumbents' overall positional extremity, which reflects electoral considerations between both the primary and general elections.

For the key independent variable of interest, the positioning of a primary challenger, I focus on only the challenger with the highest vote share. I focus on this candidate because, in certain circumstances, there is more than one primary challenger in the race. As discussed previously, I expect incumbents to be most responsive to the most threatening primary challenger in cases with more than one challenger.[12] To test this theory, I specify my independent variable as a 3-level factor variable that takes on the values of "Extreme Challenger," "Moderate Challenger," or "No Challenger" (reference category).[13] I classify challengers as extreme if they had a WEB Score greater than their parties' mean WEB Score, and moderate otherwise.[14]

[12] It should be noted, this decision only applies in cases where there is more than one challenger, which constitutes 22% of races within the scope of analysis. Of those races with more than one challenger, only 32% are instances where challenger positioning would be coded differently depending on the challenger, or 7% of all observations. Further, Appendix H shows the results are not sensitive to either (1) coding "Extreme Challenger" and "Moderate Challenger" as equal to 1 in cases where incumbent candidates are cross-pressured (see Table H1 in the appendix) or (2) changing the direction of the coding in these races (see Table H2 in the appendix).

[13] I use a factor variable instead of a continuous measure due to the fact that not all incumbents face primary challengers. With a continuous measure, there is no reliable method to place incumbents without challengers on that scale.

[14] An alternative specification would classify candidates as extreme (moderate) challengers if they were more extreme (moderate) than the incumbents' positioning in the previous election. The lack of WEB Scores for 2016 prevents the pursuit of this approach. Given I only have WEB Scores for 2018-2022, this

To provide context for where incumbents face primary challenges, Figure 3 plots the percentage of incumbents who were either challenged by an extreme candidate or a moderate candidate or were not challenged by party and election year.[15] Consistent with prior research, the majority of incumbents in both parties and election years stave off primary challengers (60%). However, there is a slight divergence in the trend across parties from 2018 to 2020 – more Democratic incumbents were challenged in 2020 (43%) than in 2018 (36%), while fewer Republican incumbents were challenged in 2020 (28%) than in 2018 (37%). This likely reflects the electoral environment and circumstances surrounding the 2020 election that presented favorable electoral circumstances to Democrats.

When considering the positioning of the challenger, overall, incumbents more often face a challenger from extreme candidates (26%) than moderate candidates (14%). However, these trends do vary by party and election year. Republican incumbents in 2018 were challenged at a higher rate by extreme candidates (22%) than moderate candidates (14%); this pattern mostly held in 2020 (18% versus 11%). For Democrats, more incumbents faced extreme primary challengers in 2018 (22%) than moderate challengers (14%). In 2020, this trend was widened with 32% of incumbents facing an extreme primary challenger while only 11% faced a moderate primary challenger.

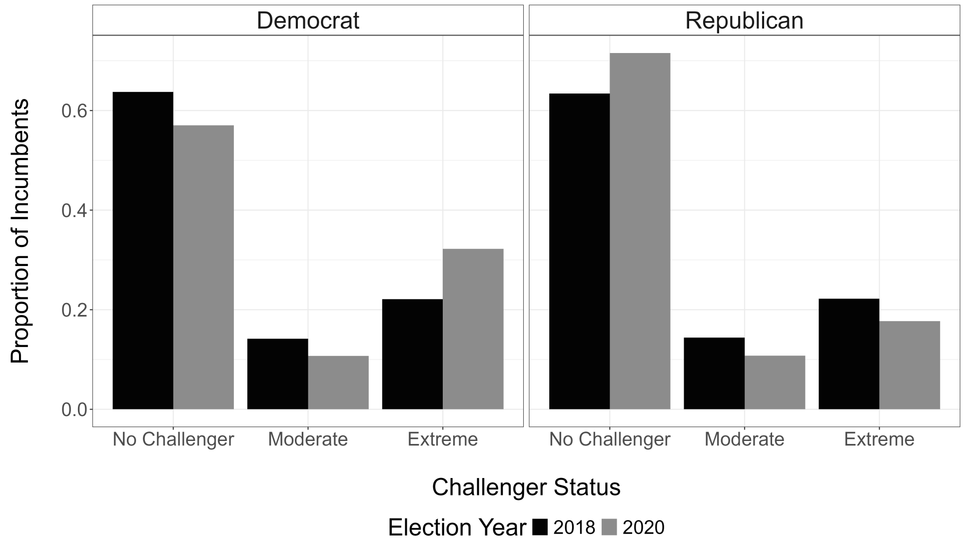
To empirically test how incumbent candidates respond to primary challengers' positioning and how these results differ by measures of positioning, I run four separate models.

would restrict my analysis to studying changes in positioning from 2020 to 2022. Over this time frame, congressional districts underwent redistricting, and the composition of districts changed. Incumbents change their rhetoric in response to changing district conditions (Kaslovsky and Kistner 2025). As a result, I could not control for district conditions between these two election cycles with incumbent fixed effects. Under that modeling strategy, changes in incumbent positioning could result from the district, not the primary challenger. However, It can be noted that in 2020, 81% of challenging candidates would be classified the same using this approach as the approach in the main results, suggesting both methods of classifying primary challengers are consistent.

[15]Incumbents where the challenger does not have a WEB Score are excluded. Figure 3 only categorizes the challenger with the highest vote share, although descriptive patterns are consistent when expanding to all challengers.

Figure 3: Incumbent Challenger Status by Party and Year

Alt Text: Bar plot showing the proportion of incumbents facing no challenger, a moderate challenger, or an extreme challenger, broken down 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.

The first two models use WEB Scores for both the dependent variable and the primary challenger classification for the independent variable. The second two models use CFScores instead of WEB Scores. 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, 33% do not have a WEB Score while 59% do not have a CFscore. While a large proportion of candidates are excluded by CFscores, WEB Scores 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.

In addition, I also include incumbent-fixed effects and year-fixed effects in columns 1

Table 4: Incumbent Positioning and Challenger Extremity

	Measurement:			
	WEB Scores		*CFScores*	
	(1)	(2)	(3)	(4)
Moderate Challenger	-0.157^{***}	-0.165^{***}	0.011	-0.001
ref: No Challenger	(0.048)	(0.042)	(0.025)	(0.024)
Extreme Challenger	0.088^{**}	0.083^{**}	0.007	-0.013
ref: No Challenger	(0.041)	(0.034)	(0.016)	(0.015)
Constant	0.826^{***}	0.746^{***}	1.129^{***}	0.888^{***}
	(0.189)	(0.027)	(0.069)	(0.018)
Observations	483	483	467	467
Incumbent Fixed Effects	✓		✓	
Incumbent Random Effects		✓		✓
Year Fixed Effects	✓	✓	✓	✓

Note: Units of analysis 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 first two columns use WEB Scores for both the independent variables and dependent variable. The second two columns use CFScores. *p<0.1; **p<0.05; ***p<0.01

and 3. Incumbent fixed effects account for unobservable time-invariant characteristics of incumbents that could affect their positioning scores, such as personal policy preferences, congressional district, primary type, or other candidate characteristics such as gender. By accounting for these factors, incumbent fixed effects also address the possibility that extreme or moderate primary challengers are more likely to emerge against certain types of incumbents (e.g., more extreme incumbents), assuming that emergence is time invariant. As a result, the coefficient can be interpreted as capturing within-incumbent variation (e.g., controlling for the positioning of the incumbent) due to changes in the status and positioning of a primary challenger. The use of TWFE is not intended to estimate a precise causal effect. Rather, TWFE accounts for time-invariant characteristics of the incumbent and unit-invariant characteristics across time that affect candidate positioning. In models 2 and 4, I also include incumbent random effects instead of incumbent fixed effects due to the low overall number of observations.

The results of all four models are presented in Table 4.[16] In all values, positive (negative) values indicate incumbents adopt more extreme (moderate) issue positions. Starting with the results in column 1, there is clear evidence incumbent candidates shift their issue positioning in response to the positioning of a primary challenger when using WEB Scores. Compared with no primary challenger, the effect of facing a moderate primary challenger is associated with incumbents moderating their positions by -0.157. To place this result in substantive terms, this is roughly similar to the difference between the average member in the Progressive Caucus and the average member in the New Democratic Coalition (-0.14), a substantively meaningful difference.

Incumbent candidates also adopt more extreme positions in response to an extreme primary challenger (relative to no primary challenger), although this effect is smaller in magnitude (0.088). When comparing the effect of an extreme primary challenger relative to a moderate primary challenger, there is evidence of incumbents shifting their issue positioning to be more extreme by 0.245 (p-value < 0.01). To place this magnitude in substantive terms, this is similar to the difference between the average Republican Main Street Partnership member and the average Republican Study Committee member (0.26). Given the differences between these ideological caucuses, this is a substantively meaningful effect and strong evidence incumbents respond to the positioning of primary challengers. Results are substantively similar when using random effects by candidate instead of fixed effects.

When conducting the same analyses using CFScores (columns 3 and 4), I find no evidence of a shift in positioning. As column 3 shows, there is not substantively or statistically significant effect of either an extreme or moderate primary challenger. These results highlight that the choice of measurement matters for the substantive conclusion. When using a measure that captures the actual issue positions members take during the campaign, incumbents respond across election years, consistent with theoretical expectations. Comparing

[16] I also run the same models with 2018, 2020, and 2022 incumbents. The results are consistent and can be found in Table H3 of Appendix H. However, the changes in congressional districts due to the 2022 redistricting cycle are not accounted for by incumbent fixed or incumbent random effects.

that result with CFScores, there is no such response. Considering the data that is underlying each measure, this result suggests while incumbents do change their campaign behavior in response to a primary challenger, campaign donors are not responsive to these changes, as incumbents CFScores do not meaningfully change.

TWFE do not control for time-variant characteristics of incumbents that could affect who challenges incumbents and pose potential endogeneity issues. For example, it could be the case that a change in political events (e.g., Covid) causes more extreme (moderate) candidates to challenge extreme (moderate) incumbents. To address the possibility of this unobserved confounder, I conduct sensitivity analysis to determine the robustness of the results in Table 4 (Cinelli and Hazlett 2020). The analysis estimates how strong an unobserved confounder would need to be, in terms of its association with the treatment and the outcome, to explain away the observed effect given a specified baseline effect size. As a baseline, I use the fixed effect for Rep. Andy Biggs (chair of the Freedom Caucus, 2019-2022) relative to Rep. Jerry Carl (reference category) and compare it to the effect of an extreme primary challenger relative to a moderate primary challenger from Table 4. The coefficient for the Andy Biggs fixed effect is 0.239. For context, this effect is similar in magnitude to the difference between the average member of the Republican Study Committee and the Main Street Partnership (diff = 0.26) and larger than the difference between the average member of the Progressive Caucus and the New Democratic Coalition (diff = 0.14). This demonstrates that the benchmark I am using is a substantively meaningful effect size as a baseline comparison.

Table 5: Sensitivity Analysis For Table 4 Model 1

Outcome: Incumbent Extremity (WEB Score)

Treatment:	Est.	S.E.	t-value	$R^2_{Y\sim D\|\mathbf{X}}$	$RV_{q=1}$	$RV_{q=1,\alpha=0.05}$
Extreme Challenger ref: Moderate Challenger	0.245	0.053	4.663	13.6%	32.6%	20.3%

Benchmark Covariate: Andy Biggs (ref: Jerry Carl); coef = 0.239

The full sensitivity analysis results are presented in Table 5. The first four columns of the table provide the coefficient, standard error, t-value, and the partial R^2 for the variable

Extreme Primary Challenger (ref: Moderate Challenger). The last two columns provide robustness values, indicating the partial R^2 needed between a confounder (with an effect of 0.239 on the outcome) on both the treatment and outcome to reduce the effect of an extreme primary challenger to zero (fifth column) or not statistically significant at the 0.05 level (sixth column). Starting with the fifth column, an unobserved confounder with an effect size of 0.239 would need to explain at least 32.6% of the residual variation in both the treatment and the outcome (i.e., have a correlation of at least 0.57) to eliminate the estimated effect of an extreme primary challenger. To render the estimate not statistically significant at the 0.05 level (sixth column), a confounder would still need to explain 20.3% of the residual variation in both variables, or a correlation of 0.45 with each. For context, the actual observed partial R^2 of challenger extremity with incumbent positioning is 13.6%, meaning the unobserved confounder would have to be substantially stronger than the effect of an extreme challenger in its relationship to the outcome, even when benchmarking against a substantively large confounder (0.24). While a TWFE design does not fully control for time-variant endogeneity concerns, these results suggest that the conclusions are reasonably robust to these concerns.

Given that I do find evidence that incumbents' campaign positioning changes, I also consider why this change in measurement is occurring, and how that relates to incumbents' issue text. As noted in the measurement construction, WEB Scores reflect both issue emphasis and issue positioning. It is important to consider how both of these components are shifting in response to a primary challenger. In Appendix I, I test for both of these possibilities. To do so, I rely data from Case and Porter (2025), which has each issue statement labeled for individual policy areas. I focus specifically on abortion, education, energy, the environment, guns, healthcare, and immigration.

I find evidence that incumbents change both what issues they discuss and their positioning on individual issue areas. In response to an extreme primary challenger (relative to a moderate primary challenger), incumbents are more likely to discuss abortion and guns; there is no effect for the other policy areas. Regarding changing positions on individual policy areas, I also find that incumbents adopt more extreme positions in response to an extreme primary challenger (relative to a moderate primary challenger) on education, energy, the environment, and healthcare. Full details of this test are included in Appendix I. The results highlight that incumbents' response to primary challengers is reflected in both changing issues discussed and changing issue positions on individual policy areas.

Conclusion

This paper provides an important contribution to the understanding of incumbent positioning in response to primary challengers and 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 improve upon the limitations of prior measurements of strategic candidate positioning. Namely, WEB Scores are conceptually distinct from previously used measures in that they capture the actual positions candidates run on during the campaign rather than approximations. In addition, WEB Scores also increase the number of candidates included within the election years of focus (2018-2022). The benefits of this new measure and the word and candidate embeddings expand the number of substantive research questions that can be answered regarding candidate positioning. Researchers can better assess how strategic campaign positioning affects election outcomes, voter behavior, representation, future legislative behavior, policy positioning of relevant groups (e.g., experienced candidates versus inexperienced candidates), among many others. Simply put, in any analysis where researchers are interested in candidates' actual campaign positioning, WEB Scores represent a comprehensive measurement for almost all candidates running for Congress post-2018. Given this, a main contribution of this project is the maintenance and distribution of WEB Scores for future election years online, along with merging variables for ease of use.

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 captures 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 exist for candidates who do not run for lower levels of 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 that future researcher consider the data-generating process and other factors that could contribute to their results and are forthcoming with these limitations.

Acknowledgements: I am grateful for helpful comments and feedback from Sarah Treul, Max Goplerud, Eric Hansen, Jeff Harden, Jacob Montgomery, Santiago Olivella, Emily Ommundsen, Rachel Porter, Jason Roberts, Tim Ryan, Brandon Stewart, as well as the editors and anonymous reviewers. 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.

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):136–159.

Aranson, Peter H. and Peter C. Ordeshook. 1972. Spatial strategies for sequential elections. In *Probability Models of Collective Decision Making*, ed. RG Niemi and HF Weisberg. Columbus, Ohio: Charles E. Merrill Publishing Company.

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.

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

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

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

Bonica, Adam. 2020. "Why Are There So Many Lawyers in Congress?" *Legislative Studies Quarterly* 45(2):253–289.

Bonica, Adam and Gary W Cox. 2017. "Ideological extremists in the US Congress: Out of step but still in office." *Available at SSRN 2970341* .

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.

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

Canes-Wrone, Brandice, David W. Brady and John F Cogan. 2002. "Out of step, out of Office: Electoral Accountability and House members' voting." *Annual Review of Political Science* 96:127–140.

Case, Colin R. and Rachel Porter. 2025. "Strategic Heterogeneity in Policy-level Positioning: Evidence from Congressional Campaigns." OSF Preprints. August 19. doi:10.31219/osf.io/zhrmv_v4.

Cinelli, Carlos and Chad Hazlett. 2020. "Making sense of sensitivity: Extending omitted variable bias." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82(1):39–67.

Coleman, James S. 1971. "Internal processes governing party positions in elections." *Public Choice* 11(1):35–60.

Cowburn, Mike and Marius Sältzer. 2025. "Partisan communication in two-stage elections: the effect of primaries on intra-campaign positional shifts in congressional elections." *Political Science Research and Methods* 13(2):392–411.

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.

Fenno, Richard F. 1978. *Home Style: House Members in Their Districts*. New York: Longman.

Gaynor, SoRelle, Kristina Miler, Pranav Goel, Alexander M Hoyle and Philip Resnik. 2025. "Express Yourself (Ideologically): Legislators' Ideal Points Across Audiences." *Journal of Politics* (Just Accepted). `https://doi.org/10.1086/736360`.

Gooch, Andrew. 2022. "Candidate repositioning, valence, and a backfire effect from criticism." *American Politics Research* 50(6):757–768.

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. 2013. *Representational Style in Congress: What Legislators Say and Why It Matters*. Cambridge University Press.

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." (Working Paper).

Hare, Christopher, David A Armstrong, Ryan Bakker, Royce Carroll and Keith T Poole. 2015. "Using Bayesian Aldrich-McKelvey Scaling to Study Citizens' Ideological Preferences and Perceptions." *American Journal of Political Science* 59(3):759–774.

Hassell, Hans JG and Neil Visalvanich. 2024. "Perceptions of Electability: Candidate (and Voter) Ideology, Race, and Gender." *Political Behavior* 46(4):2075–2098.

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):1612–1630.

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.

Kaslovsky, Jaclyn and Michael R Kistner. 2025. "Responsive rhetoric: Evidence from congressional redistricting." *Legislative Studies Quarterly* 50(3):e12473.

Le Mens, Gaël and Aina Gallego. 2025. "Positioning political texts with Large Language Models by asking and averaging." *Political Analysis* 33(3):274–282. doi: 10.1017/pan.2024.29.

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

Lee, Frances E. 2009. *Beyond Ideology: Politics, Principles, and Partisanship in the U.S. Senate*. Chicago, IL: University of Chicago Press.

Macdonald, Maggie, Megan A Brown, Joshua A Tucker and Jonathan Nagler. 2025. "To moderate, or not to moderate: Strategic domain sharing by congressional campaigns." *Electoral Studies* 95:102907.

Meinke, Scott R. 2023. *The Origins and Consequences of Congressional Party Election Agendas*. Cambridge University Press.

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.

Ornstein, Joseph T, Elise N Blasingame and Jake S Truscott. 2025. "How to train your stochastic parrot: Large language models for political texts." *Political Science Research and Methods* 13(2):264–281.

Owen, Guillermo and Bernard Grofman. 2006. "Two-stage electoral competition in two-party contests: persistent divergence of party positions." *Social Choice and Welfare* 26:547–569.

Porter, Rachel, Colin R Case and Sarah A Treul. 2025. "CampaignView, a database of policy platforms and biographical narratives for congressional candidates." *Scientific data* 12(1):1237. `https://doi.org/10.1038/s41597-025-05491-x`.

Porter, Rachel and Sarah A Treul. 2025. "Evaluating (in) experience in congressional elections." *American Journal of Political Science* 69(1):284–298.

Porter, Rachel, Sarah A Treul and Maura McDonald. 2024. "Changing the Dialogue: Descriptive Candidacies and Position Taking in Campaigns for the US House of Representatives." *The Journal of Politics* 86(2):458–474.

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.

Ringquist, Evan J. and Carl Dasse. 2004. "Lies, damned lies, and campaign promises? Environmental legislation in the 105th Congress." *Social Science Quarterly* 85(2):400–419.

Roberts, Jason M. 2007. "The Statistical Analysis of Roll-Call Data: A Cautionary Tale." *Legislative Studies Quarterly* 32(3):341–360.

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. 2023. "Embedding regression: Models for context-specific description and inference." *American Political Science Review* 117(4):1255–1274.

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.

Sulkin, Tracy. 2011. *The Legislative Legacy of Congressional Campaigns*. Cambridge University Press.

Sulkin, Tracy and Nathaniel Swigger. 2008. "Is There Truth in Advertising? Campaign Ad Images as Signals about Legislative Behavior." *Journal of Politics* 70(1):232–244.

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.

Biography: Colin R. Case is an Assistant Professor at the University of Iowa, Iowa City, IA 52242.

Online Appendix Measuring Strategic Positioning in Congressional Elections

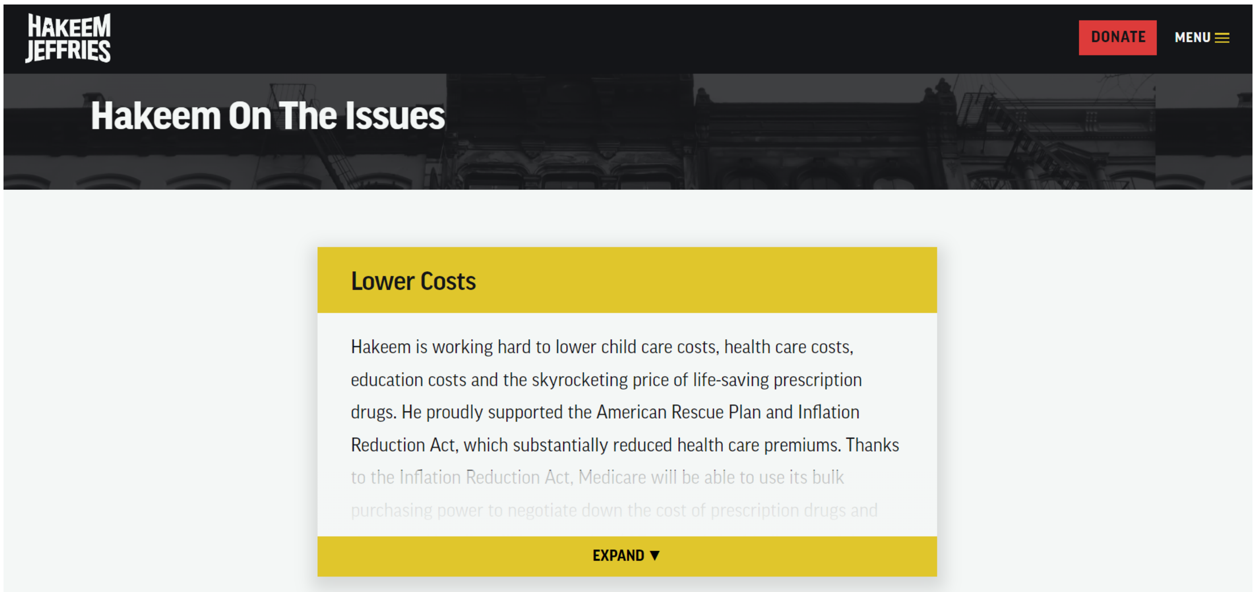
Colin R. Case

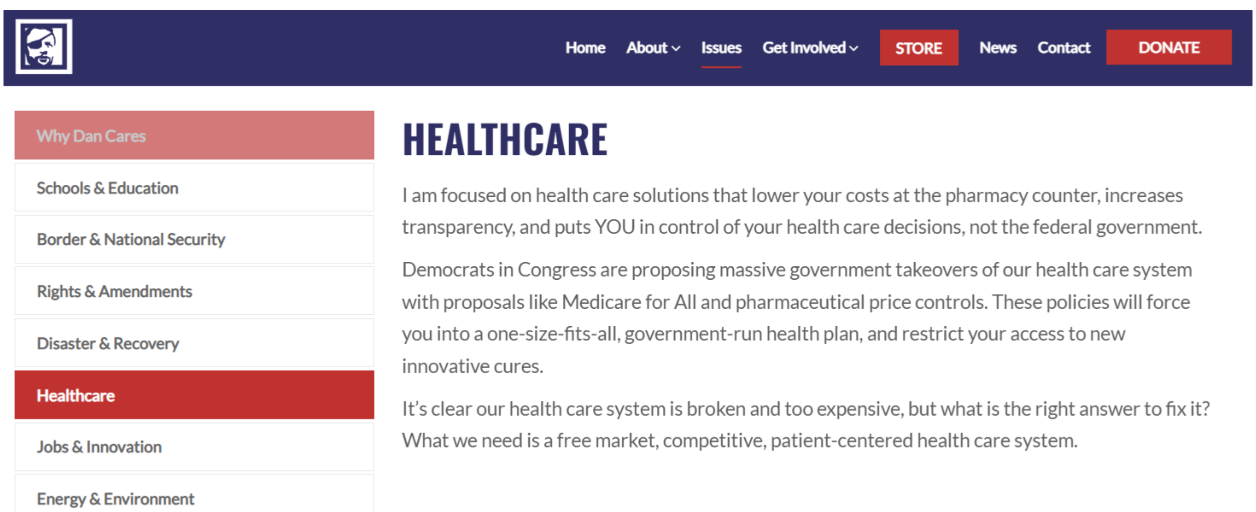
Table of Contents

Appendices

A Example Website Issue Pages

Figure A1: 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

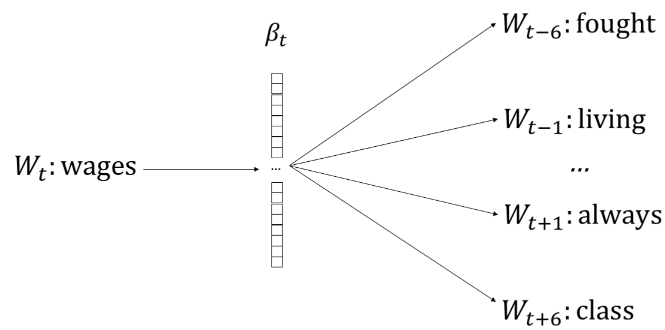
There are a number of other approaches, both supervised and unsupervised, to estimate candidate or party positioning from text. Underlying both approaches is the assumption that word usage is related to the aggregate positions that candidates take (Grimmer and Stewart 2013). One of the earliest supervised 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 "boarder" and "wall." 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.

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 B1.

Figure B1: Model Architecture with Window of 6: Word Embedding Estimation



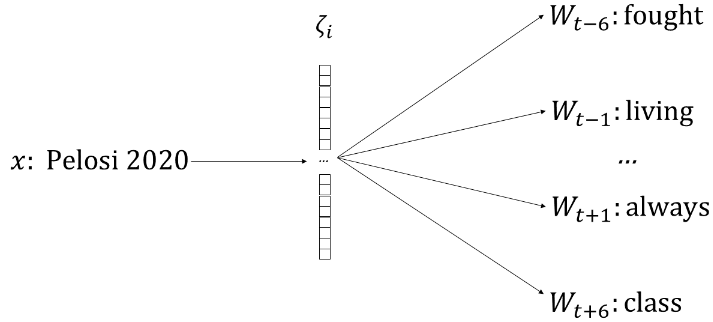
The second part of the model trains a document vector, ζ_i, for each candidate-year, i. This model architecture is the same as the first step, but instead of using a word embedding to predict words, the candidate embedding, $zeta_i$, replaces the word embedding for the target word and is used to predict 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. 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 B2. 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.[1] Removing infrequent terms improves the accuracy and performance of the models (Rodriguez and Spirling 2022).

[1] Estimates of candidate positioning are highly correlated with different cutoff thresholds (0, 5, 10, 20). See Table B1 for correlations of different hyper-parameter specifications

Figure B2: Model Architecture with Window of 6: Candidate-Year Embedding Estimation



I fit the model using parameter recommendations from Rodriguez and Spirling (2022), including a window of 6 and an embedding dimension of 300.[2] In addition, I also use pre-trained Word2Vec embeddings from the Google News corpus. These pre-trained embeddings act as initial starting weights for words in the vocabulary. The training process further fine tunes these embeddings over the text. 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 due to a limited number of documents per candidate-year occurrence and is consistent with Rheault and Cochrane (2020).

[2]Because there is no clear-cut justification for model parameters, I also fit models with various window sizes (5, 6, and 7) 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 B1 for correlation tables from different model architectures.

Model Robustness

Table B1: 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
5, 100	1	0.996	0.995	0.997	0.995	0.994	0.997	0.995	0.994
5, 200	0.996	1	0.998	0.995	0.997	0.997	0.995	0.997	0.996
5, 300	0.995	0.998	1	0.994	0.996	0.997	0.994	0.996	0.997
6, 100	0.997	0.995	0.994	1	0.996	0.995	0.997	0.995	0.994
6, 200	0.995	0.997	0.996	0.996	1	0.998	0.995	0.997	0.996
6, 300	0.994	0.997	0.997	0.995	0.998	1	0.994	0.996	0.997
7, 100	0.997	0.995	0.994	0.997	0.995	0.994	1	0.996	0.995
7, 200	0.995	0.997	0.996	0.995	0.997	0.996	0.996	1	0.997
7, 300	0.994	0.996	0.997	0.994	0.996	0.997	0.995	0.997	1

C WEB Scores: Emphasis versus Positioning

A complication of estimating policy positioning from text is understanding how changes in the text result in downstream measurement changes. In particular, there is a question on the extent to which more extreme scores are a function of either more extreme positions on a given issue (issue positions) or a larger quantity of extreme positions (issue emphasis), or both. There is a conceptual argument for measures of policy positioning reflecting both issue positions (e.g., a candidate who takes five extreme positions should measure as more extreme than a candidate who takes five moderate positions) and emphasis (e.g., a candidate who takes five extreme positions should measure as more extreme than a candidate who takes just one extreme position).

While other validation exercises in this paper demonstrate WEB Scores capture positioning, it is important to clarify the extent to which they also reflect emphasis. To do this, I create "simulated" websites using issue text from two candidates: Earl Blumenauer (OR-3, 2020) on Healthcare and Jeff Duncan (SC-3, 2022) on immigration. Blumenauer is a more ideologically extreme candidate in the Democratic Party, with one of the more liberal WEB scores among Democratic incumbents in 2020 (-1.28). Blumenauer made eleven different issue statements, almost all of which were very liberal, including statements advocating for "single-payer healthcare," "economic justice," and "abolishing ICE." Similarly, Duncan is one of the more ideologically extreme incumbent candidates in the Republican Party in 2022 (1.93). Duncan made eight issue statements, consistently running on very conservative positions such as "removing burdensome regulations," "the [fundamental] right to bear arms," and a "balanced budget amendment." Given the consistency of each candidates' extremity across statements, they provide helpful examples of how issue emphasis is reflected in WEB Scores.

For the purposes of this simulation, I use Blumenaur's issue statement on healthcare and Duncan's issue statement on immigration. In both cases, the statements clearly represent ideologically extreme positions. The full issue text is as follows:

> "Access to health care is a human right, which is why Congressman Blumenauer is fighting to expand access to quality, affordable health care to all Americans, including Medicare for All. Congressman Blumenauer is an original cosponsor of Rep. Pramila Jayapals Medicare for All bill and he is working in Congress to establish a single-payer system so that all Americans have access to quality, affordable health care. Protecting Care for Pre-Existing Conditions. In Congress, Earl has fought to defend health care for those with pre-existing conditions and has successfully fended off over 60 votes to strip away the Affordable Care Act. He opposes any bill or executive action that weakens protections for people with pre-existing conditions. Lowering Prescription Drug Prices. Drug companies have taken advantage of Americans for far too long. Congressman Blumenauer is standing up to Big Pharmas profiteering and price gouging. Hes fighting to empower Medicare and Medicaid to negotiate for lower drug prices for not only senior citizens, but all Americans. As chairman of the Ways and Means subcommittee on trade, he is fighting for lower drug prices in any trade deal negotiated."
> – Earl Blumenauer (OR-3, 2020)

> "You will not find anyone in Congress who supports securing our borders, and enforcing our immigration laws more than Jeff Duncan. Jeff has stood with President Trump has he has fought against illegal immigration, and to stop the flow of drugs and crime coming into our country. Jeff has a top rating with the pro-immigration enforcement group Numbers USA, and has been a long advocate for banning sanctuary cities, requiring mandatory E-Verify, eliminating birthright citizenship, and ending chain migration." – Jeff Duncan (SC-3, 2022)

This allows me to make a controlled demonstration of how WEB Scores reflect issue emphasis. If WEB Scores capture issue emphasis, Bluemenauer's WEB Score using all of his statements (eleven liberal positions) should be more extreme than his WEB Score for just his healthcare statement (one liberal position). Similarly, Duncan's WEB Score using all of his statements (eight conservative positions) should be more extreme than his WEB Score for just his immigration statement (one conservative position).

To make these comparisons, I take the candidates' actual issue text from the single statement above and generated a score for the candidate using only that statement. This involved re-estimating the same embedding model with the addition of two document vectors for Blumenauer and Duncan's single statements, respectively. Table C1 presents the four scores from this re-estimation. Both candidates' WEB Scores are more extreme when including all statement text than just a WEB Score generated from the single statement. This suggests WEB Scores capture emphasis in the measurement.

Table C1: WEB Scores with All Statements versus Single Statement

Candidate	All Statement WEB Score	Single Statement WEB Score
Earl Blumenauer (OR-3, 2020)	−1.27	−0.54
Jeff Duncan (SC-3, 2022)	1.93	0.86

Note: Table presents Earl Blumenauer's and Jeff Duncan's WEB Score when estimated on all of their issue statements compared with when it is estimated on just their statements on healthcare and immigration, respectively. The scores demonstrate that as candidates a larger number of extreme positions, their score becomes more extreme.

D GPT-4 and Hand-Labeled Validation

This section outlines my process for hand-labeling a subset (200) of issue statements to ensure the scores generated by GPT reflect expert-perceived scores. This task involves reading text from congressional candidates' campaign platforms and judging how much they are left-leaning (0) or right-leaning (100). For statements that take "extremely left" positions, they should be assigned a score on the lower end of the scale (e.g., 0-20). For statements that take "extremely right" positions, they should be assigned a score on the higher end of the scale (80-100). Statements that still take left- and right-leaning positions, but not as extreme as possible, should be scored in the 21-40 and 89-60 range, respectively. Statements with more centrist positions should be scored in the 41-59 range. If a statement does not have political or issue content, it should be scored as 50. Below are examples of positions that fall into various ranges from very left to very right for various policy areas. Scores for each statement are determined based on the consistency of the positions taken across the statement (e.g., a statement with extreme positions and moderate positions should be considered left/right) and the extremity of the positions taken.

Very Left (0-20)

- Abolish ICE and defund the police
- Medicare for All with no private insurance
- Green New Deal with net-zero emissions by 2030
- National rent control and universal housing guarantee
- Ban all fossil fuel extraction and fracking
- Cancel all student debt
- Federal jobs guarantee
- Decriminalize all drugs and end incarceration for drug offenses
- Ban for-profit prisons
- Abolish the Senate or Electoral College

Left (21-40)

- Public option alongside private health insurance
- Raise federal minimum wage to $15/hour
- Increased federal gun control, including assault weapons ban and red flag laws
- Pathway to citizenship for undocumented immigrants (support DACA)
- Expand union protections and collective bargaining rights
- Free community college and targeted student debt relief
- Cap insulin and prescription drug prices
- End cash bail and reduce mass incarceration
- Expand voting rights with automatic voter registration

Centrist/moderate (41-59)

- Support ACA with some reforms
- Raise minimum wage modestly, indexed to inflation
- Background checks for all gun purchases
- Border security with DACA protections
- Climate incentives without mandates
- Support school choice within public systems
- Expand mental health funding without Medicare expansion

- Compromise on abortion after first trimester
- Invest in police reform without reducing budgets

Right (60-79)
- Repeal or reduce ACA mandates
- Lower corporate taxes and reduce federal spending
- Oppose gun control; protect Second Amendment rights
- Restrict immigration; increase border enforcement
- Oppose federal climate regulations; support energy independence
- School vouchers and charter school expansion
- Oppose abortion except in exceptional circumstances (e.g., rape, incest)
- Require voter ID laws
- Increase defense spending
- Oppose labor union expansion

Very Right (80-100)
- Full repeal of ACA with no replacement
- Deport all undocumented immigrants
- Ban abortion with no exceptions
- Eliminate federal departments (e.g., Education, Energy)
- Withdraw from international climate agreements
- Arm teachers and eliminate gun-free zones
- End all federal education funding
- Flat tax or national consumption tax
- Criminalize flag burning or protests during national anthem
- Oppose all restrictions on fossil fuel production

Table D1 presents the correlations between the hand-labeled scores from this process and GPT-4-generated scores on individual statements. Consistent with work by Ornstein, Blasingame and Truscott (2025) and Le Mens and Gallego (2025), scores generated by large language models are very similar to those generated by a human rater. This provides support for using GPT-4 generated scores as a benchmark measurement. In addition, I also include correlations between hand-labeled scores on individual statements and WEB Scores, which are generated from all statement text (not individual statements). While this is an incomplete comparison, the two scores should nonetheless be related. As Table D1 shows, there is a modest relationship between human-labeled scores on a single statement and aggregated scores on all statements, providing further validation of WEB Scores.

Table D1: Correlations with Hand-Labeled Statements

Correlations with Hand-Labeled Scores			
	All Candidates	Democrats	Republicans
GPT-4 Scores	0.88	0.79	0.81
WEB Scores	0.73	0.57	0.38

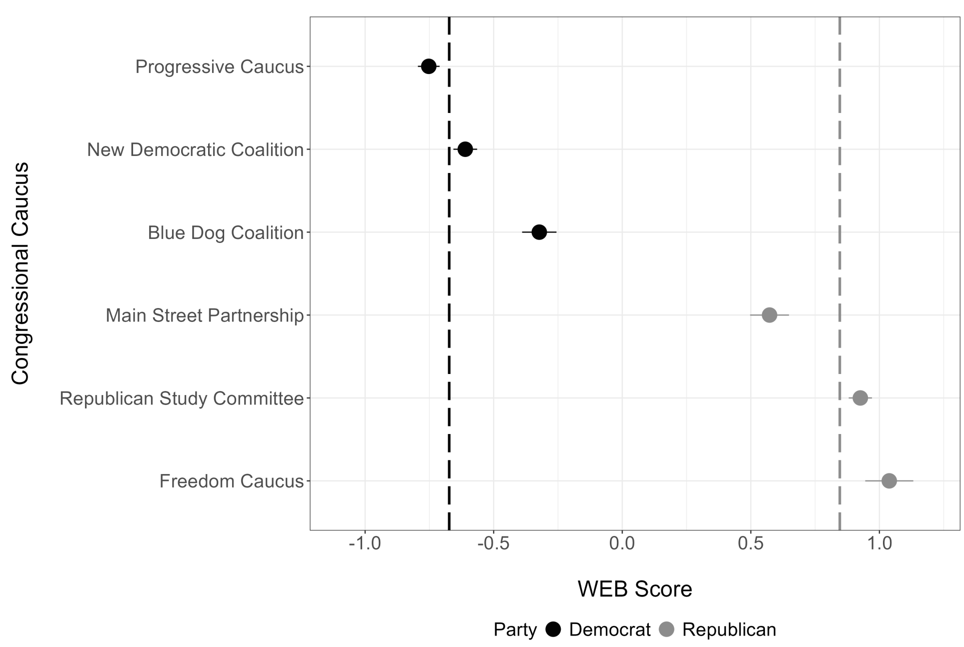
Note: Table D1 shows the correlation coefficient between hand-labeled scores and GPT-4 generated scores on individual statements as well as hand-labeled scores and WEB Scores for a sample of 200 issue statements.

E Predictive Validity Test: Caucus Membership

To provide additional external validation of WEB Scores, I also 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[3] that provided a roster of Freedom Caucus members.[4]

Figure E1: 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.

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

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

Figure E1 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 -0.75. This is less than the New Democratic Coalition, which has a mean of -0.61 (diff $= -0.14$, 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 -0.32 (diff $= -0.29$, p-value ≤ 0.001). The differences and the ideological caucus ordering match expectations and provide validity the measurement is picking up 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 0.57. Both the Republican Study Committee, with a mean of 0.93 (diff $= 0.26$, p-value ≤ 0.001), and the Freedom Caucus, with a mean of 1.04 (diff $= 0.46$, p-value ≤ 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 $= 0.11$, p-value ≤ 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.

F Content Validity Test: Embedding Relationships

In addition to validating the WEB Scores relationship with other related constructs, I also test whether or not WEB Scores capture underlying constructs in the actual embeddings that are 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:

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

In essence, this mathematical operation says the meaning of “queen” is similar to the meaning of “king” minus “man” plus “woman.” 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 and candidate embeddings exist in the same dimensional space. If WEB Scores capture variation in candidate positioning, they should also be related to certain semantic relationships that capture various positions at the candidate level. 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, the combined meaning of a liberal candidate’s statements (e.g., Alexandria Ocasio-Cortez (D-NY)), as captured by their candidate embedding, and the meaning of healthcare, as capture by the healthcare word embedding, should be closer to “universal” than the corresponding meaning for a conservative candidate (e.g., Chip Roy (R-TX)) and healthcare. This comparison can be made by adding the candidate embedding to the word embedding for healthcare and assessing the cosine similarity between the new embedding and the word embedding for “universal.”[5] It should be expected that this similarity is greater for the more liberal candidate. Specifically:

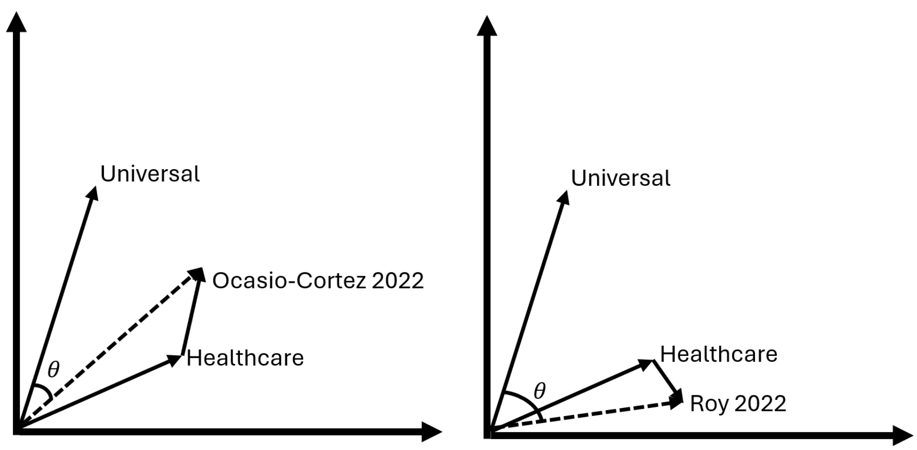
$$\begin{aligned} &cosine(vector[\text{“}healthcare\text{”}] + vector[\text{“}Ocasio - Cortez2022\text{”}], vector[\text{“}universal\text{”}]) \geq \\ &\qquad cosine(vector[\text{“}healthcare\text{”}] + vector[\text{“}Roy2022\text{”}], vector[\text{“}universal\text{”}]) \end{aligned}$$

Figure F1 provides a two-dimensional depiction of this test. As the figure shows, there is an existing relationship between the word embedding vector for “universal” and the word embedding for “healthcare.” This relationship in the vector space reflects how similar the words’ meaning is. By adding the candidate embedding for Ocasio-Cortez (left panel) and Roy (right panel), I am capturing how the language in each candidate’s statements reflects the changes in this relationship for each candidate. This is reflected by the angle, θ, between the word embedding vector for universal and the resulting vector from adding the word embedding for healthcare to the candidate embedding (depicted by the dashed line). The angle between these two vectors should be expected to be smaller for a liberal candidate (e.g., Ocasio-Cortez) than for a conservative candidate (e.g., Roy). As expected, the cosine similarity for Ocasio-Cortez is 0.28 versus 0.19 for Roy, showing the semantic similarity between “universal” and “healthcare” is closer for Ocasio-Cortez than it is for Roy. In

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

essence, this shows that the meaning of Ocasio-Cortez's issue text, relative to the word "healthcare," is closer in meaning to "universal" than it is for Roy. In the simplest terms, this test assesses whether WEB Scores capture meaningful word relationships in text that should be related to policy positioning.

Figure F1: Embedding Geometry Example



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.[6]

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 F1. Full issue statement vignettes can be found in Tables G1 and G2.[7] 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,

[6]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.

[7]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 F1: Policy Word, Policy Stances, and Keywords for Content 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

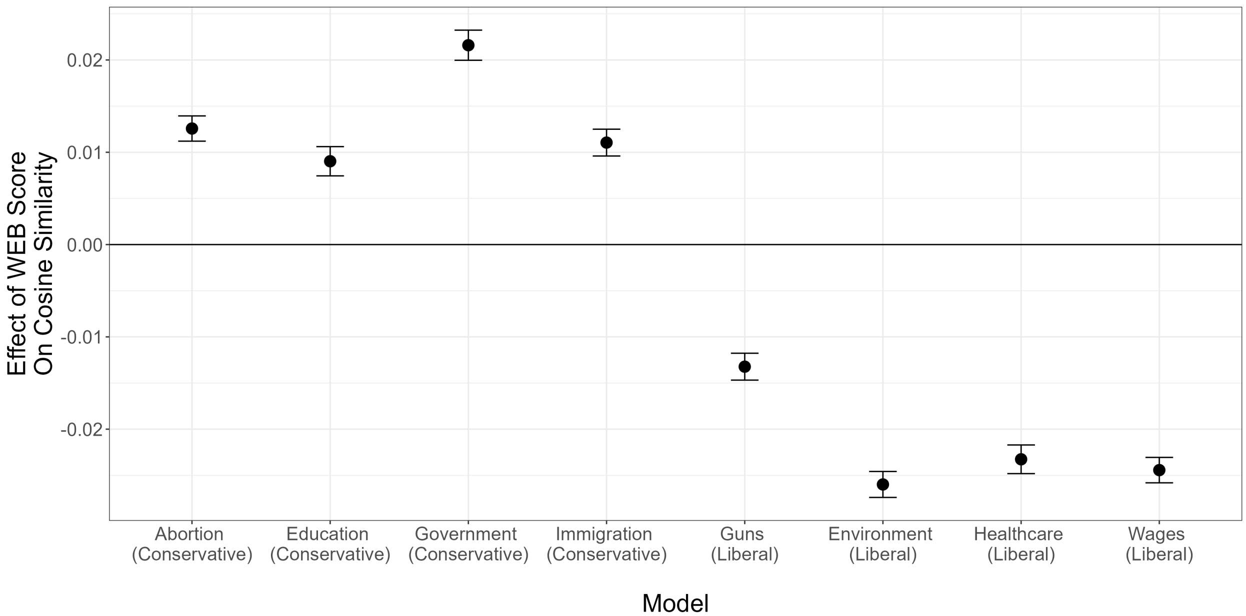
Note: Table displays the policy area (left column), the associated policy proposal advocated for by either the Justice Democrats (liberal) or the Heritage Foundation (conservative; middle column), and the keywords used in the policy proposal (right column). Full text related to policy proposal can be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

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 fit eight OLS regressions, where the dependent variable is the cosine similarity for each policy area and the independent variable is candidates' WEB Score. If WEB scores pick up on important semantic relationships related to candidate positioning, it should be expected that the coefficient is positive (negative) for conservative (liberal) policies.

Figure F2 plots the coefficient for WEB Scores from all eight models. Conservative policies are on the left side of the figure and liberal policies are on the right side of the figure. 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 positional differences across candidates in text, further validating the resulting measurement.

Figure F2: 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.

G PAC Statements

Table G1: 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 F1 are bolded in each issue statement.

Table G2: 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 F1 are bolded in each issue statement.

H Table 4 Robustness Checks

Table H1: Replication of Table 4 Column 1 Using Cross-Pressure Measure

	Measurement
	WEB Scores
Moderate Challenger Binary	-0.126^{***}
	(0.042)
Extreme Challenger Binary	0.114^{***}
	(0.038)
Constant	0.800^{***}
	(0.189)
Observations	483
Candidate Fixed Effects	✓
Year Fixed Effects	✓
Note:	*p<0.1; **p<0.05; ***p<0.01

Table H2: Replication of Table 4 Column 1 with Flipped Coding for Cross-Pressured Incumbents

	Measurement
	WEB Scores
Moderate Challenger	-0.112^{**}
ref: No Challenger	(0.044)
Extreme Challenger	0.076^{*}
ref: No Challenger	(0.042)
Constant	0.838^{***}
	(0.193)
Observations	483
Candidate Fixed Effects	✓
Year Fixed Effects	✓
Note:	*p<0.1; **p<0.05; ***p<0.01

Table H3: Replication of Table 4 with 2018, 2020, and 2022 using WEB Scores

	Measurement:			
	WEB Scores		*CFScores*	
	(1)	(2)	(3)	(4)
Moderate Challenger	-0.044	-0.058	-0.021	-0.018
ref: No Challenger	(0.040)	(0.036)	(0.018)	(0.017)
Extreme Challenger	0.019	0.034	0.001	-0.009
ref: No Challenger	(0.031)	(0.027)	(0.011)	(0.011)
Constant	-0.404^{*}	0.759^{***}	1.240^{***}	0.910^{***}
	(0.224)	(0.026)	(0.071)	(0.017)
Observations	743	743	709	709
Incumbent Fixed Effects	✓		✓	
Incumbent Random Effects		✓		✓
Year Fixed Effects	✓	✓	✓	✓

I Issue Position Change

Given I find changes in incumbent positioning in response to primary challengers, I also consider whether incumbents' response is a function of changing issue emphasis, changing issue positions on a given policy area, or a combination of both. It could be the case that incumbents change what issues they talk about (e.g., stop discussing issues that are more moderate (extreme) if they face an extreme (moderate) primary challenger). It could also be the case that candidates' actual issue stance on a given issue changes across election years, even if they talk about the issue in both years. To address both possibilities, I leverage issue statement data from Case and Porter (2025) that are coded for different policy areas (e.g., education). Issue coding for each individual statement is generated by research assistants first labeling about 9,000 different issue statements for the presence or absence of various policy areas. An ensemble machine learning classifier was trained to predict whether statements discussed a given policy area. After training, the model was used to predict policy areas discussed in the rest of the dataset.[8] I leverage this data to answer two questions: (1) are candidates more (less) likely to talk about partisan issues when challenged by an extreme (moderate) primary challenger? and (2) do candidates adopt more extreme (moderate) policy views on individual issue areas in response to an extreme (moderate) primary challenger? The answer to both these questions will address the underlying changes in behavior associated with the positioning of a primary challenger.

For this analysis, I focus on seven policy areas: abortion, education, energy, the environment, guns, healthcare, and immigration. I focus on these policy areas given that a large proportion of incumbents have an issue statement on these areas and these areas are at the center of partisan conflict in the U.S.. To assess whether or not candidates change what issues they discuss, I aggregate up from the statement level to determine whether incumbents talked about one of the seven issue areas above in a given election year. If incumbents discuss the given policy area in any issue statement, they are coded as a 1. If they do not discuss it, they are coded as a 0. To assess whether or not candidates change their issue position on a given policy area, I rely on GPT-generated labels for each individual issue statement. For each candidate, I take the average of a candidate's scores for each issue statement about a given policy area. I then subtract the party means from each candidate's score and multiply the Democrats' score by -1 so more positive (negative) values indicate more extreme (moderate) positions on that policy area. I also standardize scores to have a mean of zero and a standard deviation of 1 to provide a more interpretable measure. Importantly, candidates do not have a score if they do not talk about a given policy area in a given election year.

Similar to the analysis above, I am interested in how the positioning of a primary challenger is associated with the changes in the issues candidates discuss and their relative positioning on that issue. To conduct this analysis, I run separate models by policy area. In Table I1, my dependent variable is whether or not incumbents talked about a given policy area. My independent variable of interest is again a factor variable for moderate challenger, extreme challenger, or no primary challenger (reference category). I again control for incumbent- and year-fixed effects, so the results are within-incumbent changes. As the results demonstrate, there is mixed evidence that candidates change the issues they are discussing in response to a primary challenger. Across the various policy areas, candidates

[8] Out of sample F1-Scores for all policy areas used in this paper are above 0.8, suggesting high model performance on out-of-sample predictions.

issue uptake only changes as a function of a primary challenger for abortion (incumbents facing an extreme primary challenger are more likely to have an issue statement on abortion relative to both no challenger (p-value < 0.1) and moderate challenger (p-value < 0.05)) and guns (incumbents facing an moderate primary challenger are less likely to have an issue statement on guns relative to both no challenger (p-value < 0.05) and extreme challenger (p-value < 0.05)). For the other five issue areas, there are statistically significant effects.

Next, I turn to assessing how policy positions on a given policy area change. Here, my dependent variable is the average extremity score from GPT-generated labels on only issue statements discussing a given policy area. Positive (negative) values indicate candidates statements on a given policy area are perceived to be more extreme (moderate). As with the analysis above, I include a factor variable for the primary challenger (moderate challenger, extreme challenger, and no challenger (reference category)) as well as incumbent and year fixed effects. Importantly, if a candidate does not discuss an issue, they are not included in the analysis. Given the results are within incumbent changes, this therefore restricts the analysis coefficients to only those candidates who discussed the issue in both years. For many policy areas, as the results above illustrate, this is most candidates who discuss an issue.

The results in this analysis are presented in Table I1. In one of the seven issue areas, incumbents adopt more moderate issue positions on a given policy area relative to when there is no primary challenger that is statistically significant (education). For only one issue area (energy), there is a statistically significant effect for an extreme primary challenger relative to no primary challenger. For four of the issue areas (education, energy, environment, and healthcare), however, I find statistically significant differences in GPT-generated labels when comparing incumbents facing an extreme primary challenger relative to a moderate primary challenger. In all four instances, incumbents facing an extreme primary challenger have GPT-generated labels text related to the given policy area that is judged to be more extreme. It should be noted, these effects are modest; approximately a 0.3 standard deviation change in issue extremity. However, this should not be surprising that within incumbent changes on individual issue areas are modest changes, not drastic ones.

Table I1: Incumbent Issue Uptake and Challenger Extremity

	Issue Area:			
	Abortion	Education	Energy	Environment
	(1)	(2)	(3)	(4)
Moderate Challenger	−1.360	−42.721	−0.097	−0.100
ref: No Challenger	(2.063)	(14,632.570)	(1.279)	(1.299)
Extreme Challenger	3.415*	−0.000	1.439	0.199
ref: No Challenger	(2.039)	(1.162)	(1.408)	(1.291)
Constant	19.151	−23.566	−23.005	−22.765
	(48,196.050)	(79,462.710)	(29,232.370)	(48,196.120)
Observations	483	483	483	483
Incumbent Fixed Effects	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓

	Issue Area:		
	Guns	Healthcare	Immigration
	(5)	(6)	(7)
Moderate Challenger	−3.559**	−1.983	−0.084
ref: No Challenger	(1.607)	(1.696)	(1.001)
Extreme Challenger	1.906	0.938	−0.169
ref: No Challenger	(1.189)	(1.846)	(0.889)
Constant	19.660	−23.504	−21.397
	(29,232.430)	(48,196.060)	(29,232.480)
Observations	483	483	483
Incumbent Fixed Effects	✓	✓	✓
Year Fixed Effects	✓	✓	✓

Note: *p<0.1; **p<0.05; ***p<0.01

Table I2: Incumbent Issue-Level Change and Challenger Extremity

	Issue Area:			
	Abortion	Education	Energy	Environment
	(1)	(2)	(3)	(4)
Moderate Challenger	0.428*	−0.275*	0.076	−0.184
ref: No Challenger	(0.239)	(0.158)	(0.132)	(0.111)
Extreme Challenger	−0.109	0.091	0.398***	0.120
ref: No Challenger	(0.162)	(0.125)	(0.116)	(0.094)
Constant	0.146	−0.493	−0.442*	−0.446**
	(0.554)	(0.337)	(0.262)	(0.220)
Observations	235	338	286	279
Incumbent Fixed Effects	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓

	Issue Area:		
	Guns	Healthcare	Immigration
	(5)	(6)	(7)
Moderate Challenger	−0.211	−0.254	0.041
ref: No Challenger	(0.189)	(0.186)	(0.207)
Extreme Challenger	−0.076	0.188	−0.096
ref: No Challenger	(0.141)	(0.157)	(0.163)
Constant	−0.060	1.357***	0.752*
	(0.471)	(0.491)	(0.400)
Observations	275	418	307
Incumbent Fixed Effects	✓	✓	✓
Year Fixed Effects	✓	✓	✓

Note: *p<0.1; **p<0.05; ***p<0.01

References

Case, Colin R. and Rachel Porter. 2025. "Strategic Heterogeneity in Policy-level Positioning: Evidence from Congressional Campaigns." OSF Preprints. August 19. doi:10.31219/osf.io/zhrmv_v4.

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

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.

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.

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 Mens, Gaël and Aina Gallego. 2025. "Positioning political texts with Large Language Models by asking and averaging." *Political Analysis* 33(3):274–282. doi: 10.1017/pan.2024.29.

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

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.

Ornstein, Joseph T, Elise N Blasingame and Jake S Truscott. 2025. "How to train your stochastic parrot: Large language models for political texts." *Political Science Research and Methods* 13(2):264–281.

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.

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.

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