Public Opinion and Emphatic Legislative Speech: Evidence from an Automated Video Analysis

Oliver Rittmann* Tobias Ringwald† Dominic Nyhuis‡

Accepted for publication in the *British Journal of Political Science*

Abstract

Why do politicians sometimes deliver passionate speeches and sometimes tedious monologues? Even though the delivery is key to understanding political speech, we know little about when and why political actors choose particular delivery styles. Focusing on legislative speech, we expect legislators to deliver more emphatic speeches when their vote is aligned with the preferences of their constituents. To test this proposition, we develop and apply an automated video analysis model to speech recordings from the US House of Representatives. We match the speech emphasis with district preferences on key bills using data from the Cooperative Congressional Election Study. We find that House members who rise in opposition to a bill give more passionate speeches when public preferences are aligned with their vote. The results suggest that political actors are not only mindful of public opinion in what they say, but also in how they say it.

Keywords: Legislative Speech, Speech Delivery, Voter Preferences, Video-as-Data, US House of Representatives, Computer Vision

*Mannheim Centre for Social European Research (MZES), University of Mannheim. E-mail: oliver.rittmann@uni-mannheim.de

†Independent Researcher.

‡Institute for Political Science, Leibniz University Hanover.

Introduction

Political speech is enormously varied. While political speech is often tedious and boring, some speeches are enthralling and memorable. Whether a speech is dull or captivating depends in no small part on its delivery. Yet, even though the delivery is key in determining how a speech resonates with the public, we lack an understanding of when and why political actors make passionate appeals. In an effort to help fill this gap, this paper focuses on the nonverbal characteristics of political speech, which have rarely been the subject of systematic examination. This gap is due to measurement challenges that researchers face when analyzing nonverbal speech characteristics. Key facets of the delivery cannot be studied using the written record but can only be gauged from video footage. While political scientists have developed a rich toolbox for analyzing the textual features of political speech, the field lacks tools for capturing the nonverbal characteristics of political speech.

To overcome this limitation, we adopt methodological innovations from the field of computer vision for analyzing video data on a large scale. Building on these innovations, we develop and apply an automated video analysis model to analyze video recordings of political speeches. Specifically, we train a convolutional neural network (CNN) to analyze video footage from the US House of Representatives. The CNN is trained to detect gesturing, facial expressions, and pitch to gauge the emphasis in legislative speech.

To understand variation in speech delivery, we develop the argument that legislators adopt particular delivery styles to signal to their constituents. Scholars have long viewed political speeches as signaling tools (Mayhew, 1974). Yet, the utility of such signals depends on whether they reach their intended audience. As most political speeches go unnoticed, we argue that legislators rely on emphatic appeals to make their signaling efforts more visible. Particularly in the current media environment, it is imperative that legislators deliver good soundbites to make it past the media gatekeepers or to go viral on social media (Esser, 2008; Larsson, 2020; Negrine and Lilleker, 2002; Strömbäck, 2008). Legislators are aware of this bottleneck and they make fiery appeals when they want to

signal their positions. Whether legislators try to send such a signal depends on public opinion. When public opinion is aligned with their policy stance, we expect legislators to make an effort to highlight their position. Thus, in line with existing work on the responsiveness of political speech to public opinion (Bäck and Debus, 2018; Baumann, Debus and Müller, 2015; Hill and Hurley, 2002), we argue that legislators are not only mindful of public opinion in what they say but also in how they say it.

To assess the effect of public opinion on nonverbal speech characteristics, we link the speech emphasis with survey data. We use Multilevel Regression and Poststratification, as well as Bayesian Additive Regression Trees and Poststratification to estimate district preferences on a series of bills from the 111th to the 115th Congress (2009–2018) using data from the Cooperative Congressional Election Study (Bisbee, 2019; Warshaw and Rodden, 2012). The results provide support for the notion that legislators employ emphatic appeals to signal their policy positions when they are aligned with constituency opinion.

The findings have important implications for our understanding of legislative speech. Our study is one of few contributions that consider the nonverbal characteristics of legislative speech. In addition to showing that nonverbal speech characteristics contain valuable cues for political research, we highlight how these characteristics are shaped by strategic considerations. Among others, these findings are relevant for researchers trying to gauge the substance of political conflict from speeches (Lauderdale and Herzog, 2016; Monroe, Colaresi and Quinn, 2008; Proksch and Slapin, 2009). Incorporating the nonverbal characteristics into these efforts can generate novel insights as speech emphasis can help distinguish key policy statements from everyday speech.

Moving Beyond the Textual Features of Political Speech

Political speech is an area of intense research. Particularly the digitization of parliamentary records has helped expand our understanding of the use (Maltzman and Sigelman, 1996; Morris, 2001; Proksch and Slapin, 2012) and substance of parliamentary speech

(Hill and Hurley, 2002; Morris, 2001; Quinn et al., 2010). Despite the undeniable value of these efforts, they are subject to limitations. Efforts to categorize political speech have almost exclusively focused on textual features. While the text of speeches is sufficient for many research questions, political speech has important dimensions that are difficult to capture based on the textual features alone. Key among the characteristics that are typically disregarded in the analysis of speech is the delivery. Speeches are not generally given for the written record. They are a form of political communication where the delivery is central to their intent and effects. While some of the nontextual features may shine through in the written record, much will be lost in the transcription. For example, comparing the written record with video recordings of speeches in the Canadian House of Commons, Cochrane et al. (2022) show that emotional arousal cannot be extracted from the text alone. Succinctly put, researchers have learned a lot about what legislators say but little about how they say it.

Based on the idea that delivery matters for understanding political speech, some contributions have attempted to quantify the non-textual aspects of political speech (Banning and Coleman, 2009; Bucy, 2016; Wasike, 2019) and how they shape perceptions of the speaker (Burgoon, Birk and Pfau, 1990; Koppensteiner and Grammer, 2010; Masters and Sullivan, 1989). These efforts have been constrained by the difficulty and labor intensity of manually coding speech recordings. One promising way forward for this research is to build on the recent advances for the automated analysis of audio and video data and to apply these innovations to the ever more widely available digitized recordings of legislative speech.

A nascent literature has begun to employ these tools for researching the nonverbal characteristics of legislative speech and other recordings of political interest. While there are several studies focusing on audio recordings (Dietrich, Hayes and O'Brien, 2019; Dietrich, Enos and Sen, 2019; Knox and Lucas, 2021; Rittmann, 2024), applications studying video recordings are few and far between. For example, Dietrich (2021) uses video recordings from the US House of Representatives to analyze political polarization. Studying plenary shots, Dietrich finds that legislators have become less likely to mingle across

party lines on the House floor as polarization has gone up. Boussalis and Coan (2021) and Boussalis et al. (2021) use computer vision to extract facial expressions of candidates during televised election debates in the US and Germany and find that candidates' emotive displays affect viewers' evaluations. Finally, Neumann, Franklin Fowler and Ridout (2022) analyze politicians' body language in televised ads in the US, showing that men use more assertive gestures than women.

While existing studies constitute valuable efforts to move beyond the textual features of political speech, the current research agenda using audio and video data is fairly narrow. Due to the novelty of the data and tools for studying digitized audio and video data, the research is heavily invested in validation efforts and in exploring descriptive relationships between actor characteristics and nonverbal political behavior. What is lacking are systematic efforts to situate the new measures in conventional research programs. Indeed, the fact that previous contributions have found substantial variation in nonverbal communication underscores the need for research aimed at explaining variation in the nonverbal characteristics of legislative speech. To this end, we develop and test a theoretical account for emphatic legislative speech.

Signaling through emphatic legislative speech

The starting point for our theoretical account of emphatic legislative speech is that we conceive of speeches as signals. The signaling function of legislative speech is well established in research on political communication (e.g., Grimmer, 2013; Highton and Rocca, 2005; Hill and Hurley, 2002), yet there is a notable gap in existing accounts of speech signaling. While there is little doubt that legislators are mindful of public preferences when speaking in public, it has often remained unacknowledged how unlikely it is that such signals will reach their intended audience.

The motivating idea for this contribution is that legislators are aware that the vast majority of speech signals will go unnoticed by the public, which is why legislators can and

do make efforts to make their contributions more visible. Particularly in the current media environment, legislators can rely on a push strategy by posting their speeches on their websites or on social media. Legislators can also pursue a pull strategy by trying to create "broadcastable" moments, hoping that their contributions get picked up by traditional or social media. In practice, push and pull strategies will go hand in hand in that legislators try to create good soundbites, which they then post on social media in the hopes that their messages might go viral.

In trying to create good soundbites, legislators can rely on a host of rhetorical strategies which have been elaborated since antiquity. While captivating oratory can never be separated from the substance of what is being said, it is never a mere textual matter either. Instead, there are a great number of non-verbal characteristics that distinguish a good speech from a bad. This might cover physical aspects such as gesturing, pose, and body movement, as well as auditory features such as pace, pitch, and volume.

We expect that legislators deliberately rely on these techniques to improve the odds that their signal reaches its intended audience. Notably, we are not interested in whether a signal is actually perceived but whether legislators predictably vary their speech delivery, suggesting a deliberate use of these techniques for strategic ends. In this sense, we conceptualize speech delivery as a deliberate effort on the part of legislators rather than an unconscious indicator of legislators' true beliefs. To be sure, conceptualizing speech delivery as deliberate should not be equated with insincere. It is easily conceivable that legislators often feel strongly about an issue, resulting in a forceful delivery. Yet, the notion of deliberate emphasis suggests that, for the most part, professional politicians can choose to hide their true feelings when giving a speech if they consider doing so politically advantageous. With regards to our theoretical interest, then, we assume that legislators can choose to send a signal by way of an emphatic delivery.

There is tentative evidence that emphatic and emotional appeals are in fact more likely to be perceived by the public. Given the difficulty of systematically quantifying the emphasis in political speech (Cochrane et al., 2022), there is little direct evidence

on this question. The most direct study linking speech emphasis and media visibility is presented by Dietrich, Schultz and Jaquith (2018). Analyzing audio data from floor speeches in the US House, the authors show that more emotionally charged speeches are more likely to be broadcast and receive media coverage. Beyond the direct evidence, there are various studies on public engagement with political messages, which consistently show that greater emotional intensity predicts the success of political messages on social media (Brady et al., 2017; Heiss, Schmuck and Matthes, 2019; Nave, Shirman and Tenenboim-Weinblatt, 2018; Peeters et al., 2023), as well as showing that legislators' rhetorical skills and emotional appeals predict their visibility in traditional media (Amsalem et al., 2017; Lupacheva and Mölder, 2024; Maier and Nai, 2020; Sheafer, 2001, 2008; Sheafer and Wolfsfeld, 2004; Wolfsfeld and Sheafer, 2006).

For a first test of our new measure of legislative speech emphasis, we rely on one of the most well-established context factors in research on legislative behavior – the effect of public preferences. We expect that legislators only choose an emphatic delivery when their preferences are aligned with the preferences of their electorate. Only under conditions of alignment should we expect legislators go out of their way to try to send a signal about where they stand politically.[1]

In formulating this expectation, we are able to add nuance to research on legislative signaling. Arguably the most politically consequential signal that legislators can send is their plenary vote. While there is ample evidence that public preferences shape legislators' voting record, legislators often find themselves voting against their district, either because of strongly held personal beliefs or, more commonly, because of pressures from the party leadership, where the pressure to fall in line with the party should be especially strong

[1]It should be stressed that whether emphatic legislative speech is more visible to the public is not a precondition for the theoretical expectation to hold. It is enough for legislators to try to signal their position to the public through good soundbites when their preferences align with those of their constituents, irrespective of whether such signaling efforts are ultimately successful.

in the current climate of political polarization. Consequently, while legislators may often find themselves voting against their district, there is little reason to expect legislators to advertise that fact to their electorate by delivering an emphatic speech.

Even though there is little reason to expect legislators to emphatically highlight a vote against the district majority, one might wonder whether legislators with unpopular positions would not be better off not taking the floor at all. While trying to fly under the radar is not an unreasonable strategy, it can also prove dangerous when an unpopular vote comes to light. Therefore, legislators who find themselves voting against their district may prefer to take the floor to explain their vote. At the same time, it would rarely be wise to call attention to an unpopular stance by creating a good soundbite. We can therefore summarize that legislators whose vote is aligned with the preferences of their district are more likely to give an emphatic appeal than legislators with an unpopular stance. As an empirical matter, this expectation also helps distinguish between a deliberate and an unconscious account of speech delivery. If public preferences systematically shape legislators' speech delivery, this is unlikely to result from legislators true beliefs which accidentally shine through in their delivery.

Research Design

Are legislators more likely to deliver emphatic speeches when district preferences on a bill align with their vote? To test this proposition, we study debates on 25 pieces of legislation in the 111th–115th US House of Representatives (2009–2018). The sample was selected using four criteria. Bills were selected if they were politically salient, if public opinion data on the bills was available, if there was partisan conflict, and if public opinion toward the bill varied between congressional districts. To select the sample, we compiled a list of survey items in the Cooperative Congressional Election Survey (CCES) between 2010 and 2018 where respondents were asked to indicate their preferences on specific pieces of legislation. We then matched these questions to bills in the House of Representatives. These bills overlap to a large extent with votes that were classified as

"key votes" by *Congressional Quarterly* and cover a wide range of domestic and foreign policy issues (Ansolabehere and Jones, 2010). From this sample we discarded bills that were passed without partisan conflict[2] and bills without variation of district-level opinion. The restriction to partisan votes is plausible as voters are more likely to hold or be able to form preferences on important and controversial issues. For the same reason, signaling is a more promising strategy on important and controversial issues, as speeches on irrelevant or undisputed bills are unlikely to be observed by the public. As a practical matter, more speeches are delivered on important and partisan bills.[3] Table 1 lists the 25 bills in the resulting sample.

In the remainder of this section, we first introduce the dependent variable, the emphasis in legislative speech, and how it can be automatically gauged from plenary video recordings. Next, we discuss the estimation of district preferences on the bills as the independent variable.

---

[2]We classify votes as nonpartisan if the majority of Republican and Democratic legislators voted for or against a bill, or if more than 30% of legislators did not vote with the majority of their party.

[3]Floor access is comparatively unrestricted in the US House, both in terms of being granted speaking time and in terms of substance (Taylor, 2021). Legislative debate is generally structured by the House Committee on Rules, which specifies the rules under which a bill is brought to the floor. The Rules Committee also determines the length of the debate, which is split equally between proponents and opponents of a bill. The debate is coordinated by floor managers, usually the chair and the ranking member of the committee that reported on the bill. Therefore, while the Speaker of the House formally recognizes the individual speakers, floor access is governed by the floor managers, who allocate time to members wishing to address the assembly (Gelman and Goplerud, 2021).

Measuring Emphasis in Legislative Speech Using Automated Video Analysis

To study the emphasis in legislative speech, we analyze video recordings of key debates in the House of Representatives. The video recordings of the debates were compiled from the online archives of the House. The sample contains video recordings of all debates on the 25 pieces of legislation. We manually discard irrelevant sequences to ensure that we only analyze footage where the camera fully captures the speaker.[4] This results in 77 hours of video footage comprising 2,341 speeches by 543 legislators. Table A2 in the Online Appendix lists the debates and the pieces of legislation.

We employ computer vision to measure the speech emphasis. Specifically, we generalize manual annotations based on a set of training videos to all videos in the sample. We start by drawing an additional sample of 245 speeches by 116 legislators on 37 bills from the 115th Congress as training/test data. We manually selected 245 speeches rather than choosing them at random to ensure sufficient variation in emphasis across our training and test data. This approach allowed us to incorporate a mix of high-, mid-, and low-emphasis speeches in both datasets. The resulting videos were split into 184 training and 61 test videos. Four trained coders were tasked with annotating the emphasis in the speeches using a 7-point Likert scale, ranging from -3 (low emphasis) to $+3$ (high emphasis), for every non-overlapping two-second segment.[5]

Every video was annotated by two randomly selected coders to better judge the emphasis in the videos and to evaluate inter-rater agreement. As continuous annotation of video data is subject to different reaction times and mental processing speeds, annotations can move out of sync. Therefore, we align the annotation sequences by the two

[4]Rules for cutting the videos are documented in Online Appendix A.

[5]Coders watched full speeches and provided annotations every two seconds, such that annotations were contextually informed rather than based on isolated two-second excerpts. The coding scheme for the manual annotations is documented in Table A1 in the Online Appendix.

coders using the mean absolute error distance. This alignment shifts values by at most two seconds, i.e., by one segment. Table 2 summarizes the key values of the manually annotated data set. On average, the two annotators deviate by less than 0.5 scale points based on the mean absolute error across all two second segments in the training and test data. Additionally, we report Lin's concordance correlation coefficient and Pearson's correlation coefficient as common measures for inter-rater agreement.[6]

Predicting Speech Emphasis Using a Convolutional Neural Network

To estimate emphasis scores for speeches outside the manually annotated set, we use the training data to train a multi-modal convolutional neural network using audio and video inputs. The goal of the network is to assign emphasis scores for each two-second segment. As context information is useful for predicting the current emphasis state of a speaker, we include the surrounding two-second segments for the prediction. Thus, the model takes an input of six seconds of audio and video data for each two-second segment prediction. Before feeding the data into the network, we perform a series of preprocessing steps which we describe in Online Appendix C.

To predict the speech emphasis, we employ a convolutional neural network (CNN).[7] CNNs typically comprise two stages: feature learning and prediction. In the first stage, feature learning, the convolutional base of the model learns hierarchies of modular patterns in the input data. These features are represented in feature vectors which constitute the output of the convolutional base. In the second stage, prediction, this feature vector is fed into a second neural network which uses the features from the first stage to predict

[6]With the exception of one debate (H.R. 4760, 115th Congress), the annotated speeches are independent of those in the analysis, meaning that the training and test material does not overlap with the analysis data. For those speeches that appear in both, we relied on the human annotations in the analysis. The results remain unchanged if we exclude the speeches on H.R. 4760.

[7]See Torres and Cantú (2022) for an introduction of CNNs in the social sciences.

outcome values, in our case, emphasis scores. If trained on a large enough data set, features learned in the convolutional base are sufficiently generic to be useful for a wide variety of classification tasks. Therefore, especially in cases with small training data, pre-trained networks are commonly used for feature extraction and have proven to be highly effective (Carreira and Zisserman, 2017; Chollet and Allaire, 2018).

As our manual training data is limited, we use a pre-trained network to extract features for the video input. Specifically, we use a state-of-the art pseudo-3D-Resnet CNN (Qiu, Yao and Mei, 2017). This network is pretrained on the Kinetics data set which is commonly used for human action recognition (Kay et al., 2017). The network takes 299×299 pixel images as input and generates a 2048-dimensional feature vector. For the audio input, we use a Soundnet-like subnetwork (Aytar, Vondrick and Torralba, 2016). This network takes the audio input and generates a 512-dimensional feature vector.

After passing the video and audio inputs through these two networks in the feature learning stage, we obtain two feature vectors that summarize the video and audio inputs. In the next step, we combine both vectors and pass them to two fully connected layers. The final layer produces values in the $[-1,+1]$ range. To match the output to the original emphasis scale, we multiply the predicted values by 3 to obtain scores ranging from -3 to $+3$. Figure 1 visualizes the network architecture.

To train the model, we use a mean absolute error loss function. This function minimizes the distance between values predicted by the model and the mean emphasis scores provided by the human annotators. To prevent the neural network from overfitting, we add dropout to the fully connected layers in the second stage (Srivastava et al., 2014).[8] We use the Adam algorithm to optimize the model's parameters (Kingma and Ba, 2014).

[8]Dropout is an effective and widely used technique to prevent neural networks from overfitting. Essentially, dropout means randomly setting a number of output features of a layer in a neural network to zero during the training stage. The idea is to add noise to the output values to prevent the network from picking up on patterns that are unique to the training data.

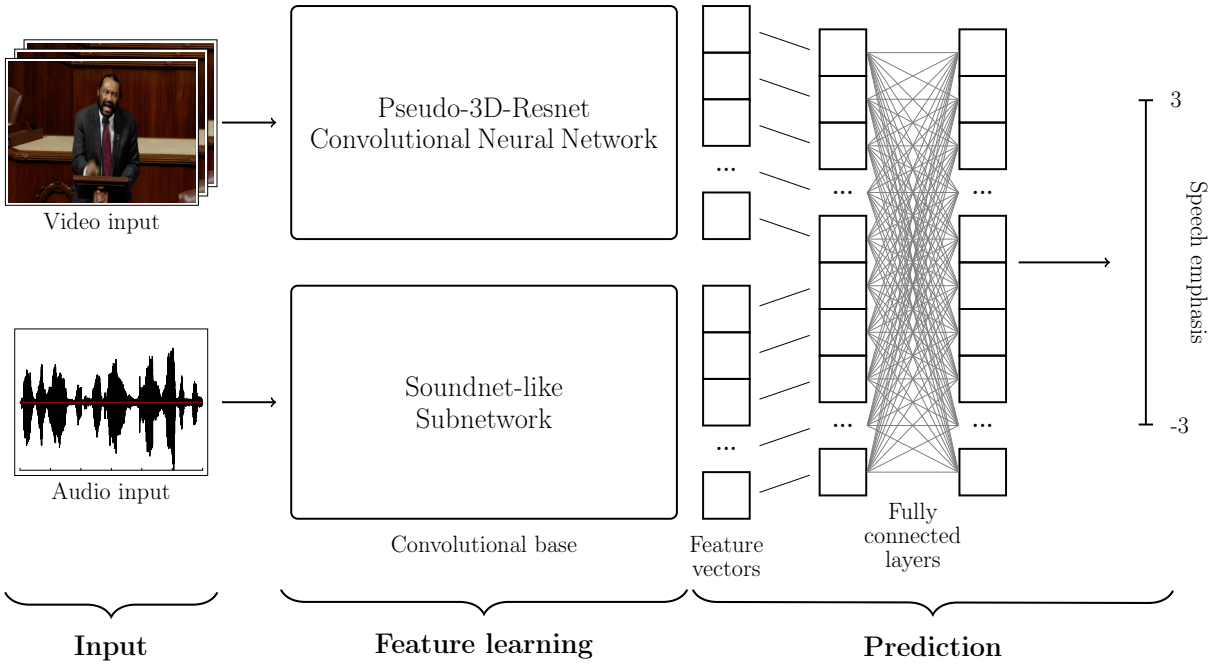


Figure 1: Convolutional neural network architecture

Applying the Model to Footage of Key Vote Debates

We apply the trained model to the video footage from all debates in our sample. The network predicts emphasis scores for each two-second sequence. For example, a one-minute speech contains 30 consecutive emphasis scores. To generate one emphasis score per speech, it is necessary to aggregate the individual scores. The simplest approach would be to calculate the average emphasis of each speech. Such an approach would ignore that speeches differ in length. This means that a multi-minute speech with 30 seconds of intense delivery would score lower than a one-minute speech with the same sequence. In line with the argument that legislators attempt to signal their issue positions by giving passionate speeches in the hopes of being amplified by traditional or social media, it is sensible to focus on shorter sequences within speeches. Legislators are aware that only short sequences of their speeches may be picked up and broadcast to the public. Therefore, it is sufficient to deliver a short, but high-intense appeal as part of a longer speech, such that short and long speeches with the same high-intense sequence should score the same. Consequently, we select the 30-second sequence with the highest within-speech average

Table 1: Summary statistics on the bills

					Emphasis	
Bill	Term	Title	Speakers	House vote	Mean	SD
HR 1	111th	Recovery and Reinvestment	86	246-183	0.33	0.60
HR 2	111th	State Children's Health Insurance	63	290-135	0.26	0.67
HR 2454	111th	Clean Energy and Security	113	219-212	-0.10	0.61
HR 3590	111th	Comprehensive Health Care Reform	91	219-212	0.15	0.60
HR 4173	111th	Financial Reform	81	223-202	0.19	0.82
HR 2965	111th	End Don't Ask Don't Tell	31	250-175	0.72	0.69
H CR 34	112th	House Budget of 2011	102	235-193	0.59	0.70
HR 2	112th	Repeal Affordable Care	231	245-189	0.39	0.69
HR 6079	112th	Repeal Affordable Care	148	244-185	0.19	0.67
HR 1938	112th	Keystone Pipeline	34	279-147	0.27	0.72
HR 1797	113th	Abortion Bill	22	228-196	0.19	0.60
HR 45	113th	Repeal Affordable Care Act	71	229-195	0.31	0.69
HR 5682	113th	Keystone Pipeline	17	252-161	0.03	0.52
HR 596	114th	Repeal Affordable Care	52	239-186	0.22	0.63
HR 3762	114th	Repeal Affordable Care	50	240-181	0.38	0.56
S 1	114th	Keystone Pipeline	19	270-152	0.29	0.50
HR 36	114th	Pain-Capable Unborn Children Protection	34	242-184	0.15	0.55
HR 3662	114th	Iran Sanctions Act	13	246-181	-0.03	0.42
HR 4760	115th	Securing America's Future	14	193-231	0.43	1.12
HR 36	115th	Pain-Capable Unborn Children Protection	43	237-189	0.06	0.63
HR 1628	115th	American Health Care	119	217-213	0.47	0.69
HR 10	115th	Financial CHOICE	67	233-186	0.17	0.73
HR 3004	115th	Kate's Law	16	257-167	-0.11	0.78
HR 3003	115th	No Sanctuary for Criminals	22	228-195	0.34	0.66
HR 1	115th	Tax Cuts and Jobs Act	91	227-203	0.56	0.73

Note: Speakers refers to the number of speakers during all debates on a bill.
House vote presents the result of the final vote on the bill.
Mean provides the average emphasis scores across all speeches on a bill,
SD is the associated standard deviation.

emphasis to score the speeches.[9] For the same reason, if a legislator delivered more than one speech on a bill, we select the speech with the highest emphasis score for the analysis. We explain this aggregation procedure in more detail in Online Appendix D.

Table 1 provides summary statistics for the resulting data. The number of legislators who delivered speeches on a bill ranges from 13 to 231. Mean emphasis scores range from -0.11 (Kate's Law) to 0.72 (End Don't Ask Don't Tell Act). Figure 2 provides additional information on the distribution of the emphasis scores across all debates, which range from

[9]As the cutoff of 30 seconds is arbitrary, we ran additional analyses where we calculate the emphasis scores for 10, 20, 40, 50, and 60-second sequences, as well as the overall average emphasis. The different choices have no effect on the substantive conclusions.

Table 2: Summary metrics for the annotated data set and model evaluation

	inter-rater baselines		naïve baselines			
	train set	test set	random guessing*	zero guessing	model prediction	30-second aggregate
Number of videos	184	61				
Number of annotated segments	12,686	3,720				
Mean absolute error	0.438	0.438	1.188 ± 0.012	0.932	0.552	0.460
Lin's concordance coefficient	0.816	0.816	-0.000 ± 0.016	–	0.764	0.837
Pearson's correlation coefficient	0.816	0.818	-0.000 ± 0.016	–	0.770	0.860

**Note:* Predictions drawn from a clipped standard normal distribution, 1000 runs.

-1.5 to 2.2. Thus, the distribution does not reach the extremes of the emphasis scale, running from -3 to $+3$. This is unsurprising as we average over 30-second segments. The distribution can be characterized as approximately normal with a mean of 0.29 and a standard deviation of 0.70.

Model Evaluation

We now turn to the evaluation of the neural network. We present the results of two validation exercises. First, we apply the trained model to the held-out test set of 61 videos and compare the model predictions with the human annotations throughout all two-second segments within those speeches. In addition, we apply our aggregation algorithm to all of these speeches using both the model predictions and the human annotations and compare the results. Table 2 shows the results of this comparison, along with the results based on random guessing (drawing values from a clipped standard normal distribution) and zero guessing (predicting an emphasis score of 0). As evaluation criteria, we compute mean absolute errors (MAE), Lin's concordance correlation coefficient (CCC) and Pearson's correlation coefficient (PCC).

Unsurprisingly, the correlations are essentially zero under random guessing. For zero guessing, the correlation is defined as zero as a constant cannot correlate with a variable. For both random and zero guessing we observe MAE values close to one standard deviation of the underlying label distribution. The neural networks achieve considerably lower MAE values. Based on the MAE metric, the machine prediction is 0.552 scale points off from

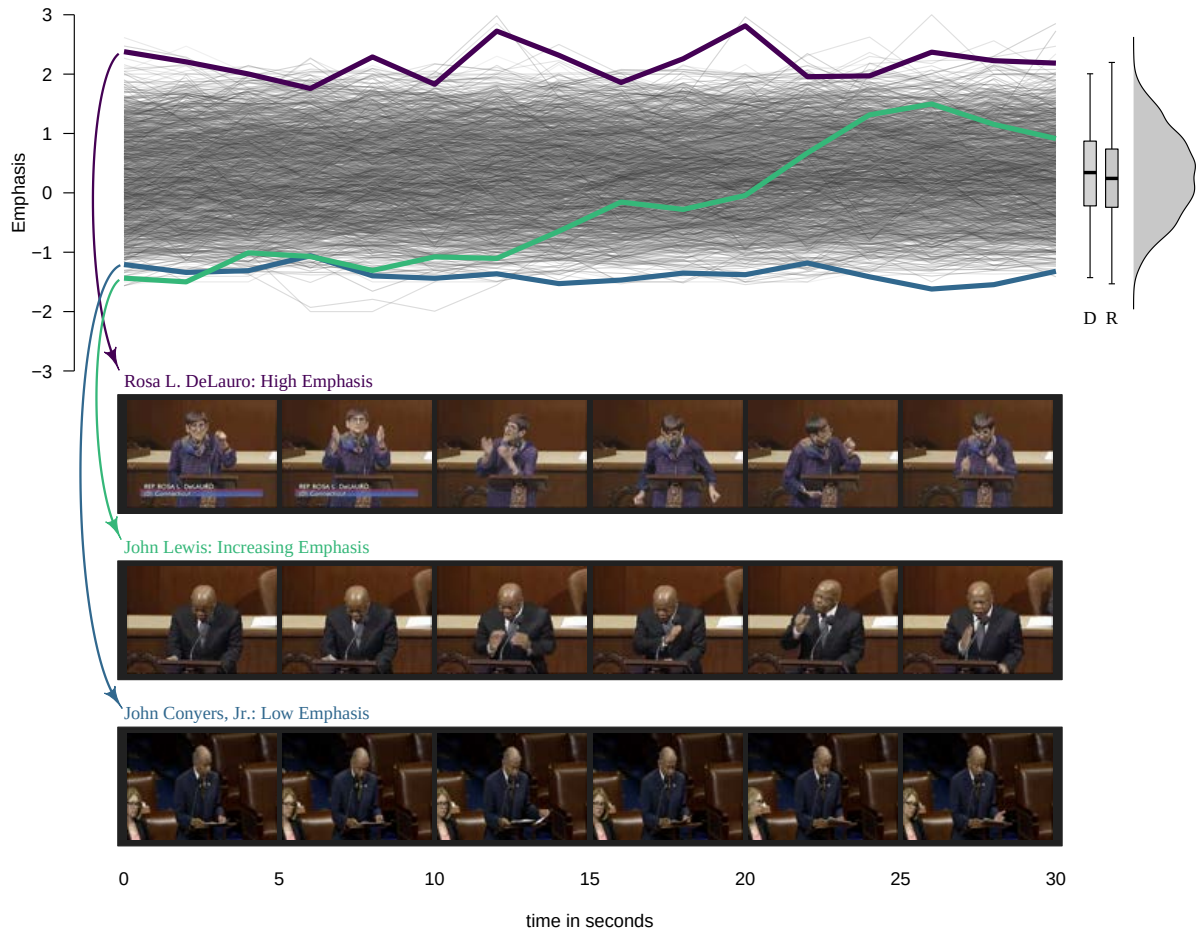


Figure 2: Visualization of the Emphasis Scores and their Distribution.

Note: Each line in the upper panel depicts the estimated speech emphasis over the course of the 30 second sequences. The three highlighted sequences depict the emphasis scores of the speeches with the highest and lowest average emphasis scores (by Rosa L. DeLauro and John Conyers, Jr.) and the speech with the highest within speech variance (by John Lewis). The video frames give an impression of how increased levels of gesturing and facial expression are linked to higher estimated emphasis scores. The density curve on the right depicts the distribution of the average emphasis scores as used in the analysis. The two boxplots represent the distributions of average emphasis scores by Democrats (D) and Republicans (R).

the human annotation when considering two-second segments. This figure is close to the human inter-rater MAE. For 30-second aggregates, the MAE decreases further to 0.460, suggesting that aggregation helps average out random noise in the data. Unlike the guessed values, the model predictions show high correlations for both the CCC and the PCC metric. As before, the correlation between the machine prediction and the human annotators is in the same range as the correlation between the human annotators. We thus conclude that the neural network reliably predicts the speech emphasis.

Our second validation is based on coders' ratings of 30- rather than two-second segments. We generated 150 speech pairs based on stratified samples from all 30-second sequences we used in the analysis.[10] For each pair, we asked two coders to indicate which of the speakers displayed a higher level of emphasis, and compared their ratings with the model ratings. Coder and model ratings are considered aligned when the speech identified as more emphatic by the coder receives a higher emphasis score from the model. If the model assigns nearly identical scores to both speeches, we would expect coder ratings to align with the model predictions in around 50% of the cases. As the difference in model emphasis scores increases, we expect coder ratings to be more likely to align with the model predictions.

Naturally, the two coders do not always agree on which of two speeches is more emphatic. Overall, they disagree on 23 of the 150 speech pairs. Panel A of figure 3 plots the probability of agreement between the coders against the difference of predicted emphasis scores by the model. It shows that the probability of the two coders agreeing on the emphasis ordering of a speech pair increases as the speeches are rated more distinct by the model.

Panel B focuses on speech pairs where both coders agree and compares their ratings

[10]Stratification is based on the predicted emphasis scores. Fifty pairs consist of random draws from above and below the median of predicted speech emphasis. Another fifty pairs are drawn from the top and bottom 25% of the emphasis distribution, and the remaining fifty pairs come from the top and bottom 10% of the predicted emphasis distribution.

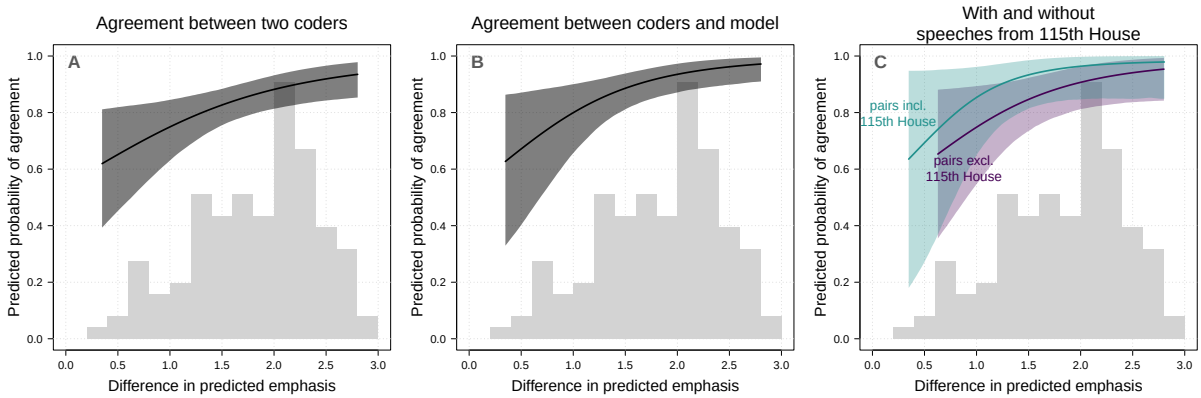


Figure 3: Model evaluation based on pairwise comparisons.

Note: Predicted probabilities and confidence intervals are based on bivariate logistic models, regressing agreement on the difference in predicted emphasis between two speeches. **Panel A** shows the predicted probability that the two coders agree on which speaker displays greater emphasis. Agreement increases as the model predicts less similar emphasis levels between the two speeches. **Panel B** is based on pairs where both coders agree on the more emphatic speech, displaying the predicted probability that their ratings align with the model predictions. Agreement between coders and the model increases as the model identifies greater differences in emphasis between the speeches. **Panel C** distinguishes between pairs that include at least one speech from the 115th legislative term and those that do not. Disagreement between the model and coder ratings is more likely for pairs without speeches from the 115th legislative term.

to the model predictions. As expected, coder and model ratings almost always align when the model predicts a substantial difference in emphasis between the two speeches. However, when the model predicts smaller differences in emphasis, coder ratings are more likely to diverge from the model predictions.

Panel C addresses the fact that the model is trained on speeches from the 115th legislative term, while the analysis includes speeches from the 111th to 115th term. These terms include speakers who were not in the trainings data, and speaking styles may have evolved, making it more difficult for the model to assess speeches from before the 115th term. To assess whether this is indeed the case, Panel C differentiates between speech pairs that include at least one speech from the 115th legislative term, and that are based on speeches from earlier terms. Indeed, the probability of alignment between coder and model ratings decreases slightly for pairs based on speeches prior to the 115th House. To address this, we include a dummy variable in our analyses, indicating whether a speech is from the 115th term or before.

While the model demonstrates satisfactory performance in the validation exercises, it is not without error. This raises the question of when and why the model makes incorrect predictions. To explore this, we conducted additional quantitative and qualitative analyses of the variation in prediction errors across speeches in our test dataset. We summarize the key insights from this analysis below and provide further details in Appendix E.

First, the model rarely predicts values above $+2$ or below -2, suggesting it may be subject to some attenuation bias. Second, our assessment of the speeches with the highest average prediction error suggests that, compared to human annotators, the model is less sensitive to hand movements occurring in front of the body of the speaker and to gestures made with a closed rather than an open hand. In contrast, the model is more responsive to large, clearly visible hand movements. This observation is plausible as open hands are better visible than closed hands and gestures stand out more clearly against the background when positioned beside the body of the speaker. Third, we observed that the model tends to overestimate the emphasis of speakers who naturally speak with a strong voice. This is understandable, as a naturally strong voice can resemble an emphatic one. The tendency may correlate with the gender of the speaker, which we account for by controlling for gender in the analysis.

Estimating District-Level Bill Preferences

The key independent variable is the extent to which legislators' votes align with constituency preferences. We draw on survey data from multiple waves of the Cooperative Congressional Election Study (CCES) to estimate district-level preferences. Each survey contains multiple questions on specific bills. Respondents are provided with the title and a short summary of the bill and are asked how they would have voted.[11] Matching each bill in our sample to a CCES item enables us to estimate district preferences toward the

[11]The question wording is: "Congress considered many important bills over the past few years. For each of the following tell us whether you support or oppose the legislation in principle."

bills.[12]

Despite the large sample size of the CCES, it is not designed to be representative of congressional district populations. Thus, simply disaggregating survey answers to estimate district-level preferences would likely yield biased estimates. To overcome this challenge, we rely on the widely used multilevel regression and poststratification (MrP) for estimating district preferences (Gelman and Little, 1997; Lax and Phillips, 2009; Warshaw and Rodden, 2012). To assess the reliability of the estimates, we complement the MrP approach with Bayesian regression trees and poststratification (BARP) (Bisbee, 2019). We employ MrP and BARP to estimate district-level preferences for the 25 bills in our sample. The estimates range from zero to one, where high values indicate high levels of support.[13]

In the next step, we match these estimates to the representatives' voting records on the 25 bills. Substantively, we are interested in the extent to which legislators' votes align with the preferences of their electorate. To compute an alignment score, we code roll call votes as one for legislators who voted in favor of a bill and zero for those who voted against it. To assess the extent to which legislators' votes align with the preferences of their electorate, we calculate the absolute difference between legislators' votes (Yes $= 1,$ No $= 0$) and the preferences of their districts and subtract this from the maximum distance, 1. We label this variable Vote-District Alignment. Values close to 1 indicate high levels of agreement between legislators and their districts, values close to 0 indicate low levels of agreement. Formally:

$$\text{Vote-District Alignment} = 1 - |\text{Yes Vote} - \text{District Preference}| \quad (1)$$

We present the distributions of the Vote-District Alignment variable by bill

[12]Table A2 in the Online Appendix reports the matches between the CCES items and the debates in our sample.

[13]Details on the estimation of district preferences are provided in Online Appendix G.

and vote choice in Figure 4. The bright density curves depict the distributions of the Vote-District Alignment for legislators who voted yes, the dark densities depict the distributions for legislators who voted no. For most bills, MrP and BARP result in similar distributions of the alignment between legislator vote and district preference. In most instances, the distribution for legislators who vote yes diverge from those who vote no. Consider the State Children's Health Insurance Act (HR 2, 111th) as an example. Although there is variation between the districts, almost all districts were fairly supportive of the bill. Thus, the Vote-District Alignment scores for legislators who voted for the bill are significantly higher than the values for legislators who voted against the bill. This also means that we observe numerous instances where legislators' votes were misaligned with their districts. Arguably, this finding can be traced back in no small part to the high levels of polarization and the resulting pressure to vote along party lines, such that legislators often find themselves between a rock and a hard place, where they either have to vote against the preferences of their constituents or risk upsetting the party leadership.

District Preferences and Signaling in Legislative Speech

To estimate the effect of district preferences on the emphasis in legislative speech, we proceed in two steps. First, we establish the link between district preferences and speech emphasis by presenting evidence from multi-level models. Next, we increase the complexity of the statistical model to explore the underlying mechanism. Specifically, we investigate whether the association is driven by differences between the delivery styles of legislators from different districts, or because legislators vary their delivery depending on the extent to which their vote is aligned with district preferences. In Online Appendix H.2, we additionally show that our results hold when using a binary measure of Vote-District Alignment.

We begin by fitting several multi-level models to estimate the effect of district opinion on speech emphasis. The dependent variable is legislators' speech emphasis. The independent variable of interest is the vote-district alignment, indicating the extent to which

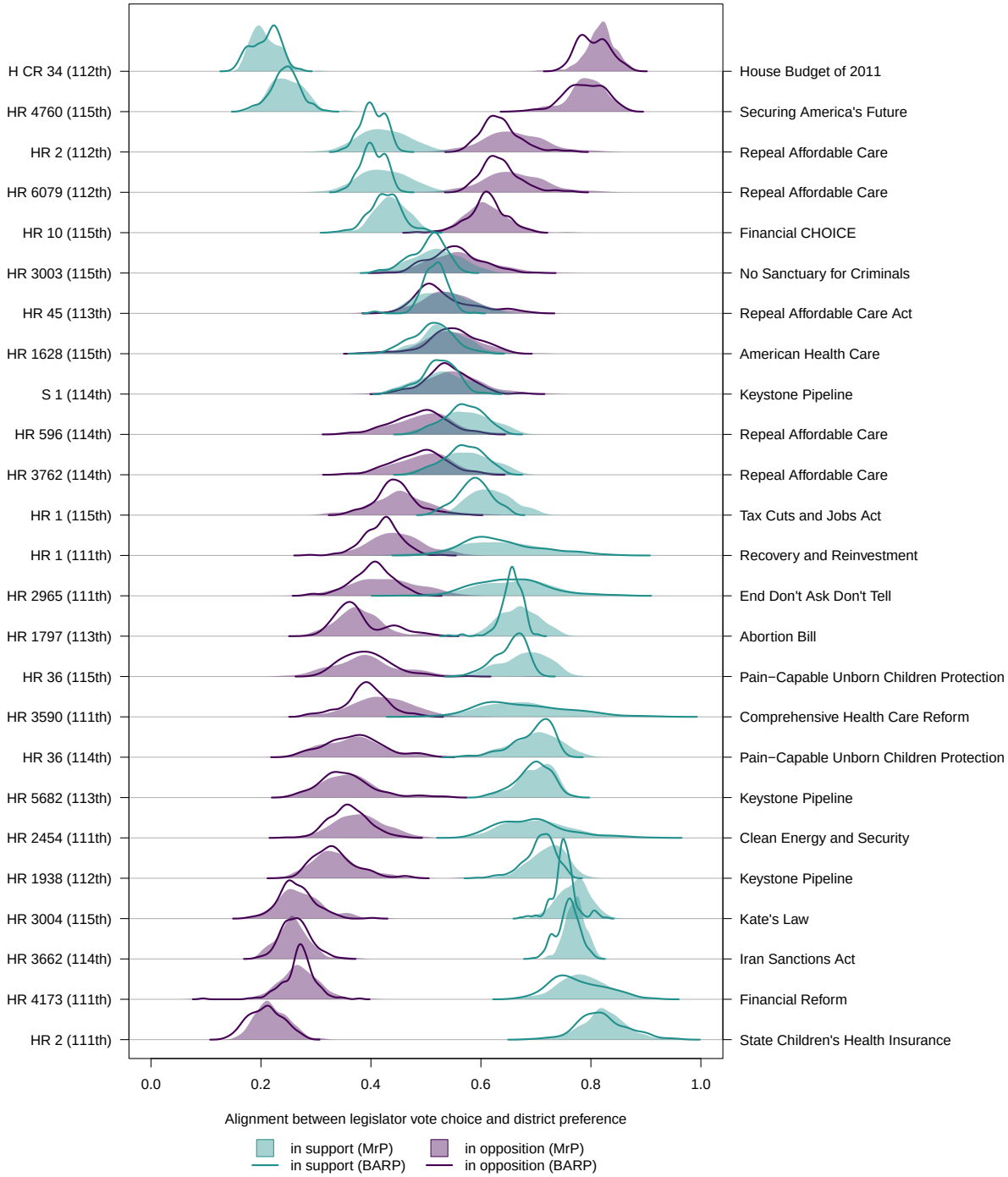


Figure 4: Distributions of alignment between legislator votes and district preferences.

legislators' floor votes are aligned with district preferences. Following the theoretical argument, we expect vote-district alignment to be positively related to speech emphasis. The model incorporates varying intercepts at the debate level to account for the hierarchical structure of the data, where speeches are nested in legislative debates on the different bills.[14]

To adjust for potential confounding, we consider six control variables. We account for legislators' party affiliation with an indicator variable for Republican legislators. As legislators who vote against the party line may face pressures not to signal this behavior, we include a binary variable that indicates whether a legislator's vote is in line with the majority of their party. We control for seniority to account for legislators' experience in delivering speeches. To account for the possibility that ideologically extreme members might deliver more emphatic speeches than moderate legislators, we include the absolute values of legislators' DW-Nominate scores (Lewis et al., 2020). Finally, we control for gender to account for the possibility that male and female legislators differ in their presentational styles, and a variable indicating whether the speech was held in the 115th term to account for potential differences in measurement error. Table 3 presents eight model specifications. Models (1) to (4) depict the results for speeches in opposition to a bill. Models (1), (3), (5), and (7) are based on MrP estimates, while models (2), (4), (6), and (8) are based on BARP estimates. Models (1), (2), (5), and (6) are baseline models without covariate adjustment, models (3), (4), (7), and (8) include control variables.

The results indicate that legislators alter their delivery style in reaction to public opinion – but only when they rise in opposition to a bill. Models (1) to (4) show a positive relation between vote-district alignment on legislator's speech emphasis, suggesting that opposing legislators deliver more emphatic speeches when their electorate is also opposed to the bill. While the magnitude of the effect remains relatively stable between the models, the substantive interpretation of the effect is not straightforward. Consider two legislators

[14]Observations can also be clustered in legislators when legislators weigh in on multiple debates. We will address this concern in the second part of the analysis.

	In opposition				In support			
	(1) MrP	(2) BARP	(3) MrP	(4) BARP	(5) MrP	(6) BARP	(7) MrP	(8) BARP
Intercept	−0.17	−0.11	−0.18	−0.15	0.21	0.27	−0.77	−0.70
	(0.14)	(0.14)	(0.24)	(0.23)	(0.19)	(0.19)	(0.33)	(0.33)
Vote-District Alignment	1.07	0.97	1.30	1.21	−0.01	−0.12	0.12	−0.01
	(0.25)	(0.26)	(0.24)	(0.28)	(0.29)	(0.30)	(0.32)	(0.32)
Controls	✗	✗	✓	✓	✗	✗	✓	✓
AIC	1592.75	1596.98	1618.53	1623.06	1716.78	1716.48	1726.01	1726.08
BIC	1611.24	1615.47	1664.74	1669.27	1735.76	1735.46	1773.47	1773.54
Log Likelihood	−792.38	−794.49	−799.26	−801.53	−854.39	−854.24	−853.00	−853.04
N	751	751	751	751	851	851	851	851
N(Debates)	25	25	25	25	25	25	25	25
Var: Debates (Intercept)	0.04	0.04	0.04	0.04	0.02	0.02	0.03	0.03
Var: Residual	0.46	0.47	0.46	0.47	0.42	0.42	0.41	0.41

The dependent variable is the level of emphasis of a legislative speech.

Table 3: Multilevel Specifications with Debate Random Effects. Parantheses report heteroskedasticity consistent wild bootstrap standard errors (Modugno and Giannerini, 2015; Loy, Steele and Korobova, 2023).

who both rise in opposition to a bill. In legislator A's district, 65% of the voters are, like legislator A, opposed to the bill. This amounts to a vote-alignment score of 0.65. In contrast, 65% of the voters in legislator B's district support the bill, which means that only 35% are aligned with their vote,[15] leading us to expect that legislator A delivers a more emphatic speech than legislator B. Based on the estimates from model (3), we expect legislator A to deliver a speech that scores 0.39 (SE = 0.07) points higher on the emphasis scale compared to legislator B.[16] This is equivalent to about 0.5 standard deviations of the distribution of speech emphasis among legislators who rise in opposition.

Turning to models (5) to (8) in Table 3, the results suggest that this finding does not generalize to legislators who deliver speeches in support of a bill. The estimated coefficients indicate that when legislators rise in support, they do not deliver their speech more or less emphatically depending on how strongly their district supports the bill. The finding that speeches in support of a bill are less affected by public opinion aligns well

[15]The 0.30 point difference of vote alignment between the two legislators amounts to about two standard deviations of the empirical distribution of the vote-alignment variable.

[16]The standard error is based on simulated first differences for a female Democrat who voted in line with her party before the 115th term. Seniority and ideology are held constant at their mean values (Ideology = 0.44, Seniority = 15.7).

with recent research which has highlighted that opposition legislators are more prone to using emotional language (Gennaro and Ash, 2022).

Before proceeding with the analysis, we should caution against interpreting the association between district preferences and speech delivery as causal. Establishing causality would require strong theoretical assumptions, especially the absence of confounding variables – assumptions we cannot confidently make given the observational nature of our study. The next section demonstrates that the association holds when considering only within-legislator variation, ensuring that the result is not driven by time-invariant confounders. While this partially addresses concerns about causality, it does not fully resolve them.

Individual Versus Macro-Level Explanation

Having shown evidence for a link between district opinion and speech emphasis among legislators who rise in opposition, we now proceed to test whether the proposed individual-level explanation is driving the findings. It is conceivable that legislators whose preferences are more consistently aligned with those of their constituents might be more likely to signal this alignment to their constituents by giving more emphatic soundbites overall. We are thus interested in differentiating between such an explanation at the legislator level from an explanation an explanation where legislators are more emphatic when their position is aligned with their district in a particular debate.

We adjust the statistical model to study the isolated effects of both explanations. To that end, we partition the total variation of vote-district alignment into two parts: Variation *within* districts and variation *between* districts. *Within* Vote-District Alignment is defined as the deviation of the vote-alignment on a specific bill from legislators' average vote-alignment, which reflects the "legislator-in-debate" level explanation. *Between* Vote-District Alignment is defined as legislators' average vote-alignment, which reflects the generic legislator level explanation.

We specify a within-between Random Effects (REWB) model (Bell, Fairbrother and

Jones, 2019; Bell and Jones, 2015) to estimate the effects of both variance components in the same model. The model is specified as follows:

$$y_{it} = \beta_0 + \beta_{1W}(x_{it} - \bar{x}_i) + \beta_{2B}\bar{x}_i + \sum_{j=1}^{k} \gamma_j z_{ji} + (\upsilon_i + \upsilon_t + \epsilon_{it}) \tag{2}$$

where y_{it} represents legislator i's emphasis in debate t on a specific bill. x_{it} is the alignment between legislator i's vote after debate t and the preference of their district. $\bar{x}_i$ is the average alignment between legislator i's votes and the preference of their district. υ_i are random intercepts for legislator i and υ_t are random intercepts for debate t. z_{ji} represent the same k individual-level control variables as before.

The model has several properties worth noting. Importantly, the independent variable of interest, VOTE-ALIGNMENT, enters the model in two forms: First, in its de-meaned form $(x_{it} - \bar{x}_i)$, i.e., the deviation of the vote-alignment from legislators' average vote-alignment. The coefficient β_{1W} represents the average *within* effect of vote-alignment, that is, the expected change in a legislator's speech emphasis caused by variation of preferences within their district. Thus, positive values of β_{1W} would constitute evidence for the individual-level explanation, making β_{1W} the main coefficient of interest. The coefficient is equivalent to individual fixed effects and is independent of differences in legislators' delivery styles. Second, the model incorporates legislators' average vote-alignment $(\bar{x}_i)$ as a covariate. The coefficient β_{2B} represents the average *between* effect of vote-alignment. This effect captures differences in emphasis between legislators with varying average levels of vote-alignment, i.e. legislators with more or less average district support for their floor votes. Thus, positive values of β_{2B} would constitute evidence for the macro-level explanation.

The results in Table 4 provide support for both the individual and the macro-level mechanism. Starting with the individual mechanism, Model (1) and (2) show that there is a positive *within* effect of vote-district alignment on speech emphasis. This result indicates that legislators who rise in opposition vary their delivery depending on how closely their

	In opposition		In support	
	(1) MrP	(2) BARP	(3) MrP	(4) BARP
Intercept	−0.29	−0.26	−0.51	−0.45
	(0.36)	(0.39)	(0.31)	(0.33)
Vote-District Alignment, *within*	0.76	0.77	−0.22	−0.27
	(0.29)	(0.30)	(0.24)	(0.24)
Vote-District Alignment, *between*	1.28	1.16	0.06	−0.04
	(0.37)	(0.41)	(0.27)	(0.29)
Controls	✓	✓	✓	✓
N (Legislators)	295	295	402	402
N (Debates)	25	25	25	25
Var: Legislators (Intercept)	0.21	0.21	0.18	0.18
Var: Debates (Intercept)	0.04	0.04	0.02	0.02
Var: Residual	0.25	0.25	0.24	0.24
AIC	1456.37	1458.26	1625.39	1625.18
BIC	1511.82	1513.72	1682.35	1682.14
Log Likelihood	−716.18	−717.13	−800.69	−800.59
N	751	751	851	851

The dependent variable is the level of emphasis of a legislative speech.

Parentheses show wild bootstrap standard errors

Table 4: Within-Between Multilevel Specifications with Legislator and Debate Random Effects. Parantheses report heteroskedasticity consistent wild bootstrap standard errors (Modugno and Giannerini, 2015; Loy, Steele and Korobova, 2023).

vote is aligned with the preference of their district. Specifically, legislators who rise in opposition deliver more emphatic speeches as their districts become increasingly hostile to the bill. Figure 5 helps to assess the size of this effect. Consider a legislator who rises in opposition in two debates on different bills. In the first debate, 50% of the voters in their district are—like them—opposed to the bill. In the second debate, 75% are opposed, making their vote more aligned with public opinion in their district.[17] Based on the estimates from model (1), we would expect the legislator to deliver a speech that scores 0.19 points higher on the emphasis scale during the second debate compared to a speech during the first debate. This is equivalent to 0.45 standard deviations of the de-meaned speech emphasis.

The results also provide evidence for the macro-level mechanism. Both models show a positive *between* effect of vote-district alignment on speech emphasis (1.28 and 1.16 de-

[17]This 25 percentage point difference is equivalent to about 2.5 standard deviations of the de-meaned vote-alignment variable ($x_{it} - \bar{x}_i$).

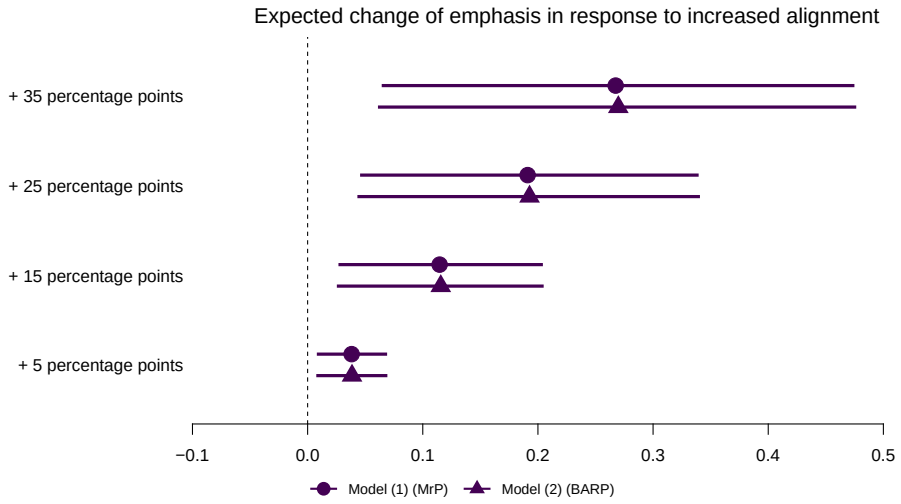


Figure 5: First Differences and 95% Confidence Intervals illustrating the expected change of speech emphasis in response to increased alignment between a legislator's No vote and public opinion in the district.

Note: Wild cluster bootstrap confidence intervals based on model (1) and model (2) in table 4. The baseline value of the de-meaned district alignment is set to -0.28 (minimum for the MrP estimates), the mean level of vote-alignment is set to the empirical mean (0.53 for MrP, 0.51 for BARP), Republican is set to zero, vote with party is set to 1, seniority is set to its mean (15.7), gender is set to zero, ideology is set to the empirical mean (0.44).

pending on the public opinion estimate). This indicates that legislators whose opposing vote constantly shows high alignment with their electorate tend to deliver more emphatic speeches compared to legislators with less support for their votes. To assess the substantive meaning of this effect, consider two otherwise similar legislators whose voters differ in their attitudes toward key pieces of legislation, where legislator A enjoys high alignment between her votes and public opinion and B does not. Suppose that the difference in vote alignment between legislators A and B amounts to 25 percentage points on average.[18] Model (1) predicts that this difference has implications for how legislators A and B present themselves on the floor: On average, legislator A would deliver speeches that score 0.32 points higher on the emphasis scale compared to legislator B. This amounts to 0.55 standard deviations of the mean emphasis score.

[18]This is equivalent to about two standard deviations of the legislators' average vote-alignment ($\bar{x}_i$).

Taken together, the results from the REWB model on legislators who rise in opposition provide evidence for both the individual-level and the macro-level mechanism. Regarding the individual-level mechanism, legislators deliver more emphatic speeches as their vote becomes more aligned with their district. At the same time, legislators whose votes constantly show high alignment with their districts deliver more emphatic speeches on average.

The results from models (3) and (4) echo the findings from the models in table 3. They show that this finding does not generalize to legislators who rise in support of a bill. The magnitudes of the estimated within and between estimates are not distinguishable from zero at conventional levels of statistical significance.

Conclusion

Automated analyses of audio and video data have begun to make their way into political science research (Dietrich, 2021; Dietrich, Enos and Sen, 2019; Dietrich, Hayes and O'Brien, 2019; Knox and Lucas, 2021, Nyhuis et al., 2021). These techniques promise to bring about significant innovations in a number of research fields by allowing scholars to make better use of the massive amounts of digitized data and to move beyond the narrow focus on digitized political text. In research on legislative speech specifically, incorporating the new tools and data sources enables the systematic study of questions beyond the substance of speeches and a greater appreciation of the nonverbal aspects of political speech.

In this paper, we have built on these nascent efforts to explain variation in the delivery of legislative speech. We have argued that legislators are not only strategic in what they say, but also in how they say it. As legislators are aware that most speeches go all but unnoticed, they make conscious decisions about when to deliver emphatic speeches in order to increase their chances of being featured in the media. Constituency preferences are a key factor in explaining such signaling in legislative speech. As actors with a singular

interest in re-election, legislators are only expected to highlight their positions when they align with the preferences of their constituents.

To assess whether the speech delivery is responsive to public opinion, we relied on automated video analyses to measure the extent to which legislators deliver emphatic speeches on 25 key bills in the 111th–115th US House of Representatives (2009–2018). The analyses have shown consistent effects of constituency opinion on speech delivery. Across different model specifications, legislators rising in opposition to a bill were found to deliver more emphatic speeches, the more their districts are opposed to the measure.

Having shown evidence for the effect of district preferences on the nonverbal characteristics of legislative speech in the US House, one might wonder to what extent the effect generalizes to other legislatures. We would argue that the effect of constituency preferences on legislative speech should be strongest where legislators can expect to benefit the most from a personal vote (cf. Carey and Shugart, 1995), such that the US House probably constitutes a most likely case for observing an effect of public preferences on speech signaling. To a somewhat lesser extent, one might expect that legislators are more mindful of their messaging in legislatures where rhetoric is more valued, such that assemblies such as the UK House of Commons is probably characterized by more emphatic appeals than its continental European counterparts (Yildirim, 2025; Osnabrügge, Hobolt and Rodon, 2021).

Despite consistent effects across different model specifications, a few limitation should be explicitly addressed. First, we argued that the Cooperative Congressional Election Study is useful for estimating district preferences on Congressional roll call votes and that attitudes toward specific bills are better suited for gauging constituency preferences and their effects on legislative speech than a general ideology measure. It should not be left unmentioned, however, that using these indicators comes at a price. As the CCES only features survey items on key congressional votes, our analysis is restricted to key debates, raising the question whether our findings generalize beyond debates on important bills. On the one hand, there is little reason to expect strategic legislators to go out of their way

to signal their position when it clashes with the preferences of their constituents. On the other hand, speeches on inconsequential bills might generally be characterized by fewer emphatic appeals, which could result in fewer differences between speeches of legislators who agree with their constituents and those who do not. Future research could shed light on the question whether our findings generalize beyond key bills by building on the present efforts and studying a broader sample of bills, while relying on a coarser measure of district ideology. Such research is greatly simplified by the promises of computer vision where trained neural networks can easily be deployed to study speeches on other bills.

Second, the observational nature of our study prevents us from making causal claims about the relationship between district preferences and emphatic speech delivery. This limitation is shared by many studies on the effect of public opinion on elite behavior, as public opinion can rarely be experimentally manipulated. However, future research might identify scenarios where naturally occurring exogenous variation in public opinion offers more credible support for the assumptions required to establish causal links between public opinion and speech delivery (cf. Hager and Hilbig, 2020). Relatedly, our research design does not offer insights into how accurately legislators assess public opinion in their district before giving a speech, leaving unanswered questions about the precise mechanism underlying our findings.

Third, future research should also try and link the textual and the nonverbal characteristics of legislative speech more closely in order to gain additional insights into legislative speech. While our study has made first steps toward such an analysis by showing how the nonverbal characteristics are tied to position taking in speeches, additional research could investigate which specific parts of speeches legislators choose to emphasize and which content features betray a high-energy delivery.

Finally, while our study is among the first to examine the determinants of emphatic legislative speech, it does not distinguish between kinesic and vocalic cues in shaping emphasis. Since kinesic cues are conveyed through video and vocalic cues through audio, future research on nonverbal components of political speech might benefit from disentan-

gling their relative contributions. In particular, methodological investigations into how audio and video modalities influence model performance could provide valuable insights into the distinct informational content carried by gestures and vocal inflection. Such analyses could help researchers prioritize modalities when computational or analytical constraints necessitate focusing on just one modality.

Overall, the study of nonverbal characteristics with emerging computer vision tools holds enormous promise for research on political speech, legislative behavior, and more. The present contribution constitutes one of the first attempts to systematically trace and explain the nonverbal characteristics of legislative speech. In line with previous research, our findings underscore that legislators are conscious and strategic in their use of legislative speech and that such strategy is not exhaustively described by the substantive aspects of legislative speech. To further refine our understanding of speech delivery, we hope that the theoretical and empirical advances presented in this contribution elicit a growing interest in the analysis of nonverbal aspects of political speech.

Supplementary Material

Online appendices are available at [reference].

Data Availability Statement

Replication data for this paper can be found at `https://doi.org/10.7910/DVN/5EWYTN`.

Acknowledgement

We thank Morten Harmening, Felix Münchow, Marie-Lou Sohnius, Sam Känner, and Caro Krömer for excellent research assistance. We are grateful to Thomas Gschwend, Lukas Stoetzer, Christian Arnold, Seo-young Silvia Kim, Patrick Kraft, Marcel Neunhoeffer, Anna Adendorf, Oke Bahnsen, Sean Carey, Franziska Quoß, Viktoriia Semenova, Christoph Steinert, and the participants of the Severyns Ravenholt Seminar in Compara-

tive Politics at the University of Washington, as well as the participants of the American Politics Research Group at the University of North Carolina at Chapel Hill for feedback on earlier versions of this manuscript.

Financial Support

This project was supported by the Deutsche Forschungsgemeinschaft and the Sonderforschungsbereich 884. Further support is gratefully acknowledged from the research alliance ForDigital.

Competing Interests

None.

References

Amsalem, Eran, Tamir Sheafer, Stefaan Walgrave and Peter John Loewen. 2017. "Media Motivation and Elite Rhetoric in Comparative Perspective." *Political Communication* 34(3):385–403.

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

Aytar, Yusuf, Carl Vondrick and Antonio Torralba. 2016. Soundnet: Learning Sound Representations from Unlabeled Video. In *Advances in Neural Information Processing Systems*. pp. 892–900.

Banning, Stephen and Renita Coleman. 2009. "Louder than Words: A Content Analysis of Presidential Candidates' Televised Nonverbal Communication." *Visual Communication Quarterly* 16(1):4–17.

Baumann, Markus, Marc Debus and Jochen Müller. 2015. "Convictions and Signals in Parliamentary Speeches: Dáil Éirann Debates on Abortion in 2001 and 2013." *Irish Political Studies* 30(2):199–219.

Bell, Andrew and Kelvyn Jones. 2015. "Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data." *Political Science Research and Methods* 3(1):133–153.

Bell, Andrew, Malcolm Fairbrother and Kelvyn Jones. 2019. "Fixed and Random Effects Models: Making an Informed Choice." *Quality & Quantity* 53(2):1051–1074.

Bisbee, James. 2019. "BARP: Improving Mister P Using Bayesian Additive Regression Trees." *American Political Science Review* 113(4):1060–1065.

Boussalis, Constantine and Travis G. Coan. 2021. "Facing the Electorate: Computational Approaches to the Study of Nonverbal Communication and Voter Impression Formation." *Political Communication* 38(1-2):75–97.

Boussalis, Constantine, Travis G. Coan, Mirya R. Holman and Stefan Müller. 2021. "Gender, Candidate Emotional Expression, and Voter Reactions During Televised Debates." *American Political Science Review* 115(4):1242–1257.

Brady, William J., Julian A. Wills, John T. Jost, Joshua A. Tucker and Jay J. Van Bavel. 2017. "Emotion Shapes the Diffusion of Moralized Content in Social Networks." *Proceedings of the National Academy of Sciences* 114(28):7313–7318.

Bucy, Erik P. 2016. "The Look of Losing, Then and Now: Nixon, Obama, and Nonverbal Indicators of Opportunity Lost." *American Behavioral Scientist* 60(14):1772–1798.

Burgoon, Judee K., Thomas Birk and Michael Pfau. 1990. "Nonverbal Behaviors, Persuasion, and Credibility." *Human Communication Research* 17(1):140–169.

Bäck, Hanna and Marc Debus. 2018. "Representing the Region on the Floor: Socioeconomic Characteristics of Electoral Districts and Legislative Speechmaking." *Parliamentary Affairs* 71(1):73–102.

Carey, John M. and Matthew Soberg Shugart. 1995. "Incentives to Cultivate a Personal Vote: A Rank Ordering of Electoral Formulas." *Electoral Studies* 14(4):417–439.

Carreira, Joao and Andrew Zisserman. 2017. Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 6299–6308.

Chollet, François and Joseph J. Allaire. 2018. *Deep Learning with R*. Manning.

Cochrane, Christopher, Ludovic Rheault, Jean-François Godbout, Tanya Whyte, Michael W.-C. Wong and Sophie Borwein. 2022. "The Automatic Analysis of Emotion in Political Speech Based on Transcripts." *Political Communication* 39(1):98–121.

Dietrich, Bryce J. 2021. "Using Motion Detection to Measure Social Polarization in the US House of Representatives." *Political Analysis* 29(2):250–259.

Dietrich, Bryce J., Dan Schultz and Tracey Jaquith. 2018. "This Floor Speech Will Be Televised: Understanding the Factors That Influence When Floor Speeches Appear on Cable Television." Paper presented at the workshop "PoliInformatics: New Data, New Methods, New Insights".

Dietrich, Bryce J., Matthew Hayes and Diana Z. O'Brien. 2019. "Pitch Perfect: Vocal Pitch and the Emotional Intensity of Congressional Speech." *American Political Science Review* 113(4):941–962.

Dietrich, Bryce J., Ryan D. Enos and Maya Sen. 2019. "Emotional Arousal Predicts Voting on the US Supreme Court." *Political Analysis* 27(2):237–243.

Esser, Frank. 2008. "Dimensions of Political News Cultures: Sound Bite and Image Bite News in France, Germany, Great Britain, and the United States." *International Journal of Press/Politics* 13(4):401–428.

Gelman, Andrew and Thomas C. Little. 1997. "Poststratification into Many Categories Using Hierarchical Logistic Regression." *Survey Methodology* 23(2):127–135.

Gelman, David and Max Goplerud. 2021. United States: Evolving Determinants of Participation in Floor Debates. In *The Politics of Legislative Debates*, ed. Hanna Bäck, Marc Debus and Jorge M. Fernandes. Oxford: Oxford University Press pp. 801–824.

Gennaro, Gloria and Elliott Ash. 2022. "Emotion and Reason in Political Language." *The Economic Journal* 132(643):1037–1059.

Grimmer, Justin. 2013. *Representational Styles in Congress: What Legislators Say and Why it Matters*. Cambridge.

Hager, Anselm and Hanno Hilbig. 2020. "Does Public Opinion Affect Political Speech?" *American Journal of Political Science* 64(4):921–937.

Heiss, Raffael, Desiree Schmuck and Jörg Matthes. 2019. "What Drives Interaction in Political Actors' Facebook Posts? Profile and Content Predictors of User Engagement and Political Actors' Reactions." *Information, Communication and Society* 22(10):1497–1513.

Highton, Benjamin and Michael S. Rocca. 2005. "Beyond the Roll-Call Arena: The Determinants of Position Taking Congress." *Political Research Quarterly* 58(2):303–316.

Hill, Kim Quaile and Patricia A. Hurley. 2002. "Symbolic Speeches in the US Senate and Their Representational Implications." *Journal of Politics* 64(1):219–231.

Kay, Will, Joao Carreira, Karen Simonyan, Brian Zhang, Chloe Hillier, Sudheendra Vijayanarasimhan, Fabio Viola, Tim Green, Trevor Back, Paul Natsev et al. 2017. "The Kinetics Human Action Video Dataset." *arXiv preprint arXiv:1705.06950* .

Kingma, Diederik P. and Jimmy Ba. 2014. "Adam: A Method for Stochastic Optimization." *arXiv preprint arXiv:1412.6980* .

Knox, Dean and Christopher Lucas. 2021. "A Dynamic Model of Speech for The Social Sciences." *American Political Science Review* 115(2):649–666.

Koppensteiner, Markus and Karl Grammer. 2010. "Motion Patterns in Political Speech and Their Influence on Personality Ratings." *Journal of Research in Personality* 44(3):374–379.

Larsson, Anders Olof. 2020. "Winning and Losing on Social Media: Comparing Viral Political Posts Across Platforms." *Convergence* 26(3):639–57.

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

Lax, Jeffrey R. and Justin H. Phillips. 2009. "How Should We Estimate Public Opinion in the States?" *American Journal of Political Science* 53(1):107–121.

Lewis, Jeffrey B., Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin and Luke Sonnet. 2020. "Voteview: Congressional Roll-Call Votes Database (2020).".
URL: *https://voteview. com*

Loy, Adam, Spenser Steele and J. Korobova. 2023. "lmeresampler: Bootstrap Methods for Nested Linear Mixed-Effects Models." *R package version 0.2.4* .

Lupacheva, Tatiana and Martin Mölder. 2024. "A Place to Speak and Be Heard? Parliamentary Speech and Media Attention in Estonia, 2011-2019." *Legislative Studies Quarterly* 49(4):905–924.

Maier, Jürgen and Alessandro Nai. 2020. "Roaring Candidates in the Spotlight: Campaign Negativity, Emotions, and Media Coverage in 107 National Elections." *International Journal of Press/Politics* 25(4):576–606.

Maltzman, Forrest and Lee Sigelman. 1996. "The Politics of Talk: Unconstrained Floor Time in the US House of Representatives." *The Journal of Politics* 58(3):819–830.

Masters, Roger D. and Denis G. Sullivan. 1989. "Nonverbal Displays and Political Leadership in France and the United States." *Political Behavior* 11(2):123–156.

Mayhew, David R. 1974. *Congress: The Electoral Connection*. Yale University Press.

Modugno, Lucia and Simone Giannerini. 2015. "The Wild Bootstrap for Multilevel Models." *Communications in Statistics: Theory and Methods* 44(22):4812–4825.

Monroe, Burt L., Michael P. Colaresi and Kevin M. Quinn. 2008. "Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict." *Political Analysis* 16(4):372–403.

Morris, Jonathan S. 2001. "Reexamining the Politics of Talk: Partisan Rhetoric in the 104th House." *Legislative Studies Quarterly* 26(1):101–121.

Nave, Nir Noon, Limor Shirman and Keren Tenenboim-Weinblatt. 2018. "Talking it Personally: Features of Successful Political Posts on Facebook." *Social Media and Society* 4(3).

Negrine, Ralph and Darren G. Lilleker. 2002. "The Professionalization of Political Communication: Continuities and Change in Media Practices." *European Journal of Communication* 17(3):305–323.

Neumann, Markus, Erika Franklin Fowler and Travis N. Ridout. 2022. "Body Language and Gender Stereotypes in Campaign Video." *Computational Communication Research* 4(1):254–274.

Nyhuis, Dominic, Tobias Ringwald, Oliver Rittmann, Thomas Gschwend and Rainer Stiefelhagen. 2021. Automated Video Analysis for Social Science Research. In *Handbook of Computational Social Science, Volume 2*. Routledge pp. 386–398.

Osnabrügge, Moritz, Sara B. Hobolt and Toni Rodon. 2021. "Playing to the Gallery: Emotive Rhetoric in Parliaments." *American Political Science Review* 115(3):885–899.

Peeters, Jeroen, Michael Opgenhaffen, Tim Kreutz and Peter van Aelst. 2023. "Understanding the Online Relationship Between Politicians and Citizens: A Study on the User Engagement of Politicians' Facebook Posts in Election and Routine Periods." *Journal of Information Technology and Politics* 20(1):44–59.

Proksch, Sven-Oliver and Jonathan B. Slapin. 2009. "Position Taking in European Parliament Speeches." *British Journal of Political Science* 40(3):587–611.

Proksch, Sven-Oliver and Jonathan B. Slapin. 2012. "Institutional Foundations of Legislative Speech." *American Journal of Political Science* 56(3):520–537.

Qiu, Zhaofan, Ting Yao and Tao Mei. 2017. Learning Spatio-Temporal Representation with Pseudo-3D Residual Networks. In *Proceedings of the IEEE International Conference on Computer Vision*. pp. 5533–5541.

Quinn, Kevin M., Burt L. Monroe, Michael Colaresi, Michael H. Crespin and Dragomir R. Radev. 2010. "How to Analyze Political Attention with Minimal Assumptions and Costs." *American Journal of Political Science* 54(1):209–228.

Rittmann, Oliver. 2024. "Legislators' Emotional Engagement with Women's Issues: Gendered Patterns of Vocal Pitch in the German Bundestag." *British Journal of Political Science* 54(3):937–945.

Sheafer, Tamir. 2001. "Charismatic Skill and Media Legitimacy: An Actor-Centered Approach to Understanding the Political Communication Competition." *Communication Research* 28(6):711–736.

Sheafer, Tamir. 2008. "Charismatic Communication Skill, Media Legitimacy, and Electoral Success." *Journal of Political Marketing* 7(1):1–24.

Sheafer, Tamir and Gadi Wolfsfeld. 2004. "Production Assets, News Opportunities and Publicity for Legislators: A Study of Israeli Knesset Members." *Legislative Studies Quarterly* 29(4):611–630.

Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever and Ruslan Salakhutdinov. 2014. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting." *The Journal of Machine Learning Research* 15(1):1929–1958.

Strömbäck, Jesper. 2008. "Four Phases of Mediatization: An Analysis of the Mediatization of Politics." *International Journal of Press/Politics* 13(3):228–246.

Taylor, Andrew J. 2021. Legislative Speech in Presidential Systems. In *The Politics of Legislative Speech*, ed. Hanna Bäck, Marc Debus and Jorge M. Fernandes. Oxford: Oxford University Press pp. 51–71.

Torres, Michelle and Francisco Cantú. 2022. "Learning to See: Convolutional Neural Networks for the Analysis of Social Science Data." *Political Analysis* 30(1):113–131.

Warshaw, Christopher and Jonathan Rodden. 2012. "How Should We Measure District-Level Public Opinion on Individual Issues?" *The Journal of Politics* 74(1):203–219.

Wasike, Ben. 2019. "Gender, Nonverbal Communication, and Televised Debates: A Case Study Analysis of Clinton and Trump's Nonverbal Language during the 2016 Town Hall Debate." *International Journal of Communication* 13:251–276.

Wolfsfeld, Gadi and Tamir Sheafer. 2006. "Competing Actors and the Construction of Political News: The Contest Over Waves in Israel." *Political Communication* 23(3):333–354.

Yildirim, Tevfik Murat. 2025. "Emotions in the Aisles: Unpacking the Use of Emotive Language in the UK House of Commons." *European Journal of Political Research* 64(2):943–959.

Online Appendices:

Public Opinion and Emphatic Legislative Speech: Evidence from an Automated Video Analysis

Oliver Rittmann
University of Mannheim

Tobias Ringwald
Independent Researcher

Dominic Nyhuis
Leibniz University Hanover

A Rules for Cutting Videos

This section documents the rules for manually cutting the videos into sequences for measuring emphasis in legislative speech. A speech begins after the Speaker recognizes a member who wants to deliver a speech ("the gentleman/gentlewoman is recognized for ... minutes"). A speech ends after the speaker yields back his or her time. Formal phrases such as "I yield myself such time as I may consume" are not considered part of a speech. The starting point is set when the camera fully captures the speaker for the first time. This can happen after the legislator begins delivering their speech. If a speech starts at "02:47:15" but the camera only focuses on the legislator between "02:47:27" and "02:47:28", the sequence starting at "02:47:28" is used for studying the speech emphasis. The legislator must be fully captured by the camera during the entire time frame between the start and end time. If a speech is interrupted, e.g., because a floor manager grants additional time to a speaker or because the camera does not focus on the speaker, the sequence ends and a new sequence begins when the speaker is on screen again. For each sequence, we document whether it represents a new speech or whether it constitutes the continuation of an ongoing speech. If a speech consists of multiple sequences, all sequences belonging to the same speech are merged.

B Manual Video Annotation

		Posture/gestures	Audio
Very high	+3	Very strong gestures, high level of body movement	High-paced speech, screaming, yelling
	+2	Open posture, strong gestures, high level of body movement	Fast-paced speech, loud voice
	+1	Gaze forward, open posture, notable gestures and body movement	Elevated speech pace, slightly raised voice
Medium	0	Frequent gaze forward, open posture, some gestures and body movements	Fluent, conversational speech pace, conversational pitch,
	−1	Frequent gaze forward open posture, few gestures or body movements	Fluent, but slow speech pace, little emphasis in speech
	−2	Gaze down, reading, closed posture, little body movement, little use of hands	Monotone, low voice
Very low	−3	Gaze down, reading, closed posture, no body movement	Notable pauses in speech, low voice

Table A1: Manual speech emphasis coding scheme

C Video and Audio Pre-Processing

To pre-process the video input, we resize, normalize, and randomly crop the input images. Random cropping helps prevent the model from overfitting to the training data and increases the model generalizability as the input data is slightly modified every training epoch (Taylor and Nitschke, 2018). To ensure reproducibility, we use center cropping for the test footage. The final visual input are images with a size of 299×299 pixels.

For the audio input, we extract 20 Mel-frequency cepstral coefficients (MFCCs) from the raw audio data and feed them into the network (cf. Huang, Acero and Hon, 2001, ch. 6).

D Aggregation of Emphasis Scores

This section illustrates how we aggregate the output of the video analysis model into a single emphasis score per speech. Our model generates an emphasis score for every 2-second segment, resulting in a time series of scores for the entire speech. The upper panel of Figure A1 visualizes this time series for a sample speech.

To derive a single emphasis score, we identify the 30-second segment with the highest average emphasis. We achieve this by computing a moving average over a 30-second window, as shown in the lower panel of Figure A1. The solid black line represents the moving average, while the horizontal gray lines indicate the corresponding 30-second segments. The highest moving average value determines the target segment, i.e., the one with the greatest emphasis.

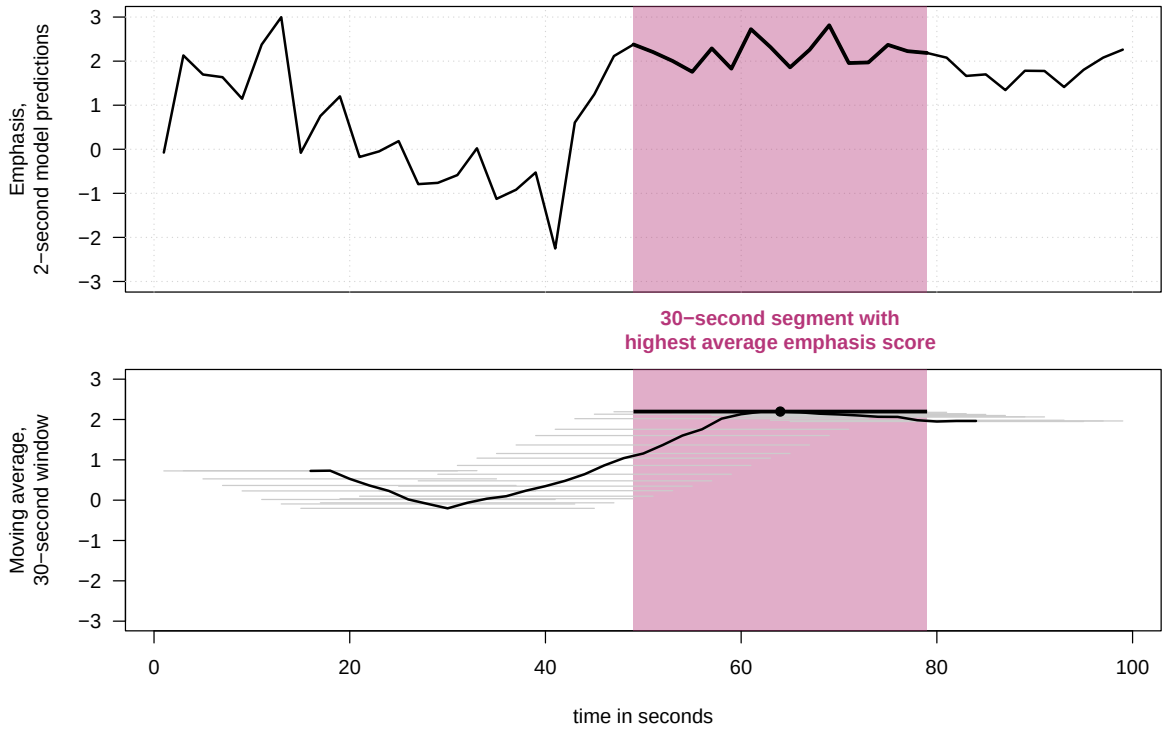


Figure A1: Illustration of emphasis aggregation procedure.

E Qualitative Assessment of Prediction Error

While the model performs satisfactorily in validation exercises, it is not without error. This raises the question: When and why does the model make incorrect predictions? In this section, we present findings from a qualitative analysis examining variation in prediction errors across speeches in our test dataset. Specifically, we re-watched recordings of speeches where the model predictions deviated most and least from human annotations. While this analysis does not provide definitive evidence, it offers valuable insights into possible non-random errors in the model and helps researchers theorize about their impact on downstream analyses.

Our examination yielded three key insights. First, the model rarely predicts the most extreme values: There are almost no speeches that reach an emphasis level above 2 or below -2, even though such values exist in the annotated data. Thus, there is some attenuation bias. We do not consider this as problematic, as long as the correlation between model predictions and human annotations hold.

Second, the qualitative viewing of the speeches with the highest average prediction error suggests that, compared to human annotators, the model is less sensitive to hand movements occurring in front of the body of the speaker and to gestures made with a closed rather than an open hand. Figure A2 illustrates this point. It shows frames from two sample speeches. In the examples, the model underestimates the annotator emphasis assessment when the speakers use gestures in front of their bodies with a closed hand (Panels A.1 and B.1). Conversely, when speakers gesture next to their bodies with open rather than closed hands, the model predictions increase while the annotator assessments remain stable (Panel A.2).

The model also appears to be more sensitive than human annotators to gestures involving both hands rather than one (Panels A.3 and B.2). This discrepancy likely arises because the model is more responsive to large, clearly visible hand movements, while human annotators additionally infer emphasis from more subtle gestures that still convey

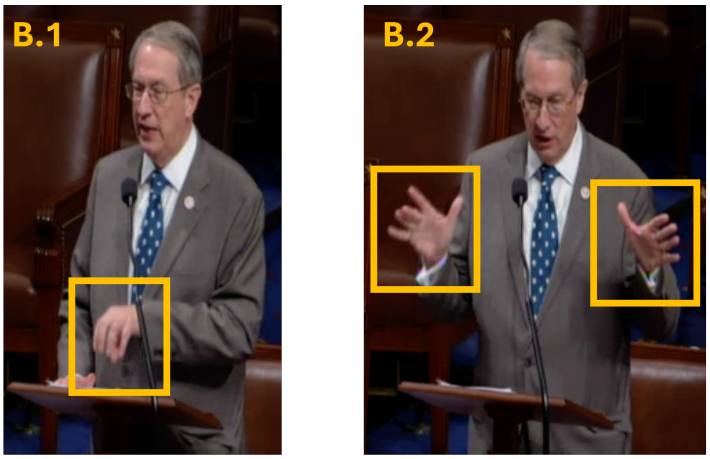
intentionality. These observations are plausible: open hands are more visible than closed hands and stand out more clearly against the background when positioned beside the body of the speaker. This is particularly evident in Panel A.1.

Third, the model tends to overestimate the emphasis of speakers who naturally speak with a strong voice. This is understandable, as a naturally strong voice can resemble an emphatic one. This tendency may correlate with the gender of the speaker, as it is plausible that speakers with strong voices are more likely to be male – underlining the importance of controlling for gender in the analysis.

Target = 1.00
Prediction = −0.53
Bias = −1.53

Target = 0.50
Prediction = 0.18
Bias = −0.32

Target = 0.50
Prediction = 0.40
Bias = −0.10



Target = 1.00
Prediction = −0.81
Bias = −1.81

Target = 1.00
Prediction = 0.44
Bias = −0.56

Figure A2: Selected frames to illustrate the varying sensitivity of human coders and the prediction model to different forms of hand movement. While the model infers higher level of emphasis from hand movement with open hands next to the body of the speaker (Panels A.2, A.3, and B.2), human annotators are equally sensitive to more subtle but intentional hand movements in front of the body of the speaker (Panel A.1 and B.1).

F Matching Legislative Debates, Bills, and CCES Items

Table A2: CCES items matched with bills, legislative debates and House roll calls

Wave	Item	Bill	Congress	Title	House Vote (Yea-Nay)	Date
2010	CC332A	HR 1	111th	American Recovery and Reinvestment Act	246-183	01/28/2009
				—Conference report	246-183	02/13/2009
2010	CC332B	HR 2	111th	State Children's Health Insurance Program	289-136	01/14/2009
				—Agree to Senate Amendment	290-135	02/04/2009
2010	CC332C	HR 2454	111th	American Clean Energy and Security Act	219-212	06/26/2009
2010	CC332D	HR 3590	111th	Comprehensive Health Reform Act		
				—Agree to Senate Amendment (& HR 4872)	219-212	03/21/2010
2010	CC332F	HR 4173	111th	Financial Reform Bill		12/09/2009
				—Unfinished business		12/10/2009
				—On passage	223-202	12/11/2009
2010	CC332G	HR 2965	111th	End Don't Ask, Don't Tell		
				—Agree to Senate Amendment	250-175	12/15/2010
2012	CC332A	H CR 34	112th	House Budget (of 2011)		04/14/2011
				—Unfinished business	235-193	04/15/2011
2012	CC332G	HR 2	112th	Repeal Affordable Care Act		01/18/2011
				—Remaining five hours of debate	245-189	01/19/2011
2012	CC332G	HR 6079	112th	Repeal Affordable Care Act		07/10/2012
				—Unfinished business	244-185	07/11/2012
2012	CC332H	HR 1938	112th	Keystone Pipeline	279-147	07/26/2011
2013	CC332A	HR 1797	113th	Abortion Bill	228-196	06/18/2013
2013	CC332C	HR 45	113th	Repeal Affordable Care Act	229-195	05/16/2013
2013	CC332D	HR 5682	113th	Keystone Pipeline	252-161	11/13/2014
				—Agree to Conference Report	251-166	01/29/2014
2015	CC15_327A	HR 596	114th	Repeal Affordable Care	239-186	02/03/2015
2015	CC15_327A	HR 3762	114th	Repeal Affordable Care	240-189	10/23/2015
				—Agree to Senate Amendment	240-189	01/06/2016
2015	CC15_327B	S 1	114th	Keystone Pipeline	270-152	02/11/2015
2015	CC15_322c	HR 36	114th	Pain-Capable Unborn Child Protection Act	242-148	05/13/2015
2016	CC16_351G	HR 3662	114th	Iran Sanctions Act	246-181	01/13/2016
2017	CC17_331_5	HR 4760	115th	Securing America's Future Act of 2018	193-234	06/21/2018
2017	CC17_332c	HR 36	115th	Pain-Capable Unborn Child Protection Act	237-189	10/03/2017
2017	CC17_340C	HR 1628	115th	American Health Care Act		03/24/2017
				—Resumed debate	217-213	05/04/2017
2017	CC17_340D	HR 10	115th	Financial CHOICE Act	233-186	06/08/2017
2017	CC17_340E	HR 3004	115th	Kate's Law	257-167	06/29/2017
2017	CC17_340G	HR 3003	115th	No Sanctuary for Criminals Act	228-195	06/29/2017
2017	CC18_326	HR 1	115th	Tax Cuts and Jobs Act		11/15/2017
				—Resumed debate	224-201	11/16/2017
				—Conference report	227-203	12/19/2017

G Measuring District Preferences

The key independent variable is the distance between legislators' roll call vote and their districts' mean preference on a bill. We employ two approaches to estimate district preferences on the bills listed in Table 1: Multilevel regression with poststratification (MrP) (Gelman and Little, 1997; Warshaw and Rodden, 2012) and Bayesian additive regression trees with poststratification (BARP) (Bisbee, 2019). BARP follows the logic of MrP but replaces the multilevel model with a Bayesian additive regression tree model. For both approaches, two types of data are needed: (1) To model individual bill preferences, survey data on respondents' preferences, their district, and demographic characteristics. For this, we draw on data from multiple waves of the Cooperative Congressional Election Study (CCES). The CCES includes questions on preferences toward specific bills (dependent variables of the multilevel model), as well as information on respondents' congressional district and demographics (independent variables). (2) To estimate district preferences through poststratification, we use census data with information on the joint distribution of demographic and geographic information of voters in the congressional districts.[19] This data is provided by the US Census Bureau. Specifically, we draw on the American Community Survey (ACS). All district-level data sets are listed in Table A3.

G.1 Multilevel Regression with Poststratification

The estimation procedure closely follows Warshaw and Rodden (2012). We measure individual preferences toward specific bills using the "roll-call" items in the CCES. Table A2 reports how we match individual bills to specific CCES items. To model individual responses, we employ a multilevel model including respondents' race (white, black, hispanic, other), gender, education (measured in four categories), and congressional district. At the district level, we include respondents' state, the median household income in the district,

[19]Data on total population estimates was retrieved from Manson et al. (2020).

percentage of veterans, the natural log of the population density,[20] and the share of same-sex marriages. At the state level, we further include presidential vote shares in the previous presidential election and the percentage of evangelical Protestants and Mormons. The latter data is based on the Religious Congregations and Membership Study (Jones et al., 2002). We incorporate this information in the hierarchical model as follows:

$$Pr(y_i = 1) = logit^{-1}(\alpha_0 + \alpha^{\text{race}}_{r[i]} + \alpha^{\text{gender}}_{g[i]} + \alpha^{\text{educ}}_{e[i]} + \alpha^{\text{district}}_{d[i]}) \tag{3}$$

where

$$\alpha^{\text{race}}_r \sim N(0, \sigma^2_{\text{race}}), \text{ for r = 1, ..., 4} \tag{4}$$

$$\alpha^{\text{gender}}_g \sim N(0, \sigma^2_{\text{gender}}) \tag{5}$$

$$\alpha^{\text{educ}}_e \sim N(0, \sigma^2_{\text{educ}}), \text{ for e = 1, ..., 4} \tag{6}$$

We model district effects as a function of the state, its median income, the share of veterans in the district, the natural log of the population density, and the share of same-sex marriages in the district:

$$\begin{aligned} \alpha^{\text{district}}_d \sim & N(\alpha^{\text{state}}_{s[d]} + \gamma^{\text{inc.}} \times \text{income}_d + \gamma^{\text{vet.}} \times \text{veterans}_d + \\ & \gamma^{\ln(\text{popdensity})} \times \ln(\text{popdensity})_d + \\ & \gamma^{\text{samesex}} \times \text{samesex}_d, \sigma^2_{\text{district}}), \\ & \text{for d = 1, ..., 435} \end{aligned} \tag{7}$$

The state effects are modeled as a function of the state-level presidential vote shares and the percentage of Evangelical and Mormon residents in the state:

[20] We use shapefiles and population estimates to estimate the population density. Shapefiles are provided by Lewis et al. (2013) and the US Census Bureau.

$$
\begin{aligned}
\alpha_s^{\text{state}} \sim & N(\gamma_0 + \\
& \gamma^{\text{presvote}} \times \text{presvote}_s + \\
& \gamma^{\text{relig.}} \times \text{religion}_s, \sigma^2_{\text{state}}), \\
& \text{for s} = 1, ..., 50
\end{aligned}
\tag{8}
$$

We begin by estimating the model for each key vote. Next, we build the poststratification data set with one row for every possible combination of predictors in each district, along with the district and state-level information. The poststratification data set contains the share of residents in each district that exhibit the various combinations of individual characteristics. Based on the model predictions for individuals with each combination of factors, the district preferences are estimated as a linear combination of the predicted preferences for individuals with the different combinations weighted by the true share of residents in the district with the respective combination. This yields an estimate of the district preference toward all bills in the sample ranging from zero to one, where low values indicate opposition to the bill and high values indicate support.

G.2 Bayesian Additive Regression Trees with Poststratification

BARP was introduced by Bisbee (2019). BARP relies on the same logic as MrP but replaces the multilevel model with a fully nonparametric regularization technique, Bayesian additive regression trees (BART). Due to its nonparametric character, BARP allows for deep interactions between covariates without requiring the researcher to specify these functional forms when setting up the model. Thus, BARP is less vulnerable to model misspecification compared to MrP. We use the same data as before.[21] For estimation, we

[21]Note that because BARP relies on nonparamteric regularization, instead of taking the natural log of district-level population densities, we include population density as an untransformed variable.

rely on the R-Packages `BARP` (Bisbee, 2019) and `bartXViz`.

Estimation Results

Figure A3 depicts the distributions of the resulting district preference estimates by bill. Figure A4 shows the bivariate distribution of the MrP and BARP estimates.

Table A3: District-level poststratification data

CCES wave	Legislative term	Year (Census)	Survey	Description	Dataset
2010	111th	2010	ACS 1 year estimates	Sex by educational attainment, race	C15002 (H, B, I)
2010	111th	2010	ACS 1 year estimates	Income in the past 12 months	S1903
2010	111th	2010	ACS 1 year estimates	Veteran Status	S2101
2010	111th	2010	ACS 1 year estimates	Household and Families	S1101
2010	111th	2010	ACS 1 year estimates	Total Population Estimates	B01003
2012	112th	2010	ACS 1 year estimates	Sex by educational attainment, race	C15002 (H, B, I)
2012	112th	2010	ACS 1 year estimates	Income in the past 12 months	S1903
2012	112th	2010	ACS 1 year estimates	Veteran Status	S2101
2012	112th	2010	ACS 1 year estimates	Household and Families	S1101
2012	112th	2012	ACS 1 year estimates	Total Population Estimates	B01003
2013	113th	2010	ACS 1 year estimates	Sex by educational attainment, race	C15002 (H, B, I)
2013	113th	2014	ACS 1 year estimates	Income in the past 12 months	S1903
2013	113th	2014	ACS 1 year estimates	Veteran Status	S2101
2013	113th	2014	ACS 1 year estimates	Household and Families	S1101
2013	112th	2013	ACS 1 year estimates	Total Population Estimates	B01003
2014	113th	2014	ACS 1 year estimates	Sex by educational attainment, race	C15002 (H, B, I)
2014	113th	2014	ACS 1 year estimates	Income in the past 12 months	S1903
2014	113th	2014	ACS 1 year estimates	Veteran Status	S2101
2014	113th	2014	ACS 1 year estimates	Household and Families	S1101
2014	112th	2014	ACS 1 year estimates	Total Population Estimates	B01003
2015	114th	2015	ACS 1 year estimates	Sex by educational attainment, race	C15002 (H, B, I)
2015	114th	2014	ACS 1 year estimates	Income in the past 12 months	S1903
2015	114th	2014	ACS 1 year estimates	Veteran Status	S2101
2015	114th	2014	ACS 1 year estimates	Household and Families	S1101
2015	112th	2015	ACS 1 year estimates	Total Population Estimates	B01003
2016	114th	2015	ACS 1 year estimates	Sex by educational attainment, race	C15002 (H, B, I)
2016	114th	2014	ACS 1 year estimates	Income in the past 12 months	S1903
2016	114th	2014	ACS 1 year estimates	Veteran Status	S2101
2016	114th	2014	ACS 1 year estimates	Household and Families	S1101
2016	112th	2016	ACS 1 year estimates	Total Population Estimates	B01003
2017	115th	2017	ACS 1 year estimates	Sex by educational attainment, race	C15002 (H, B, I)
2017	115th	2017	ACS 1 year estimates	Income in the past 12 months	S1903
2017	115th	2017	ACS 1 year estimates	Veteran Status	S2101
2017	115th	2017	ACS 1 year estimates	Household and Families	S1101
2017	112th	2017	ACS 1 year estimates	Total Population Estimates	B01003

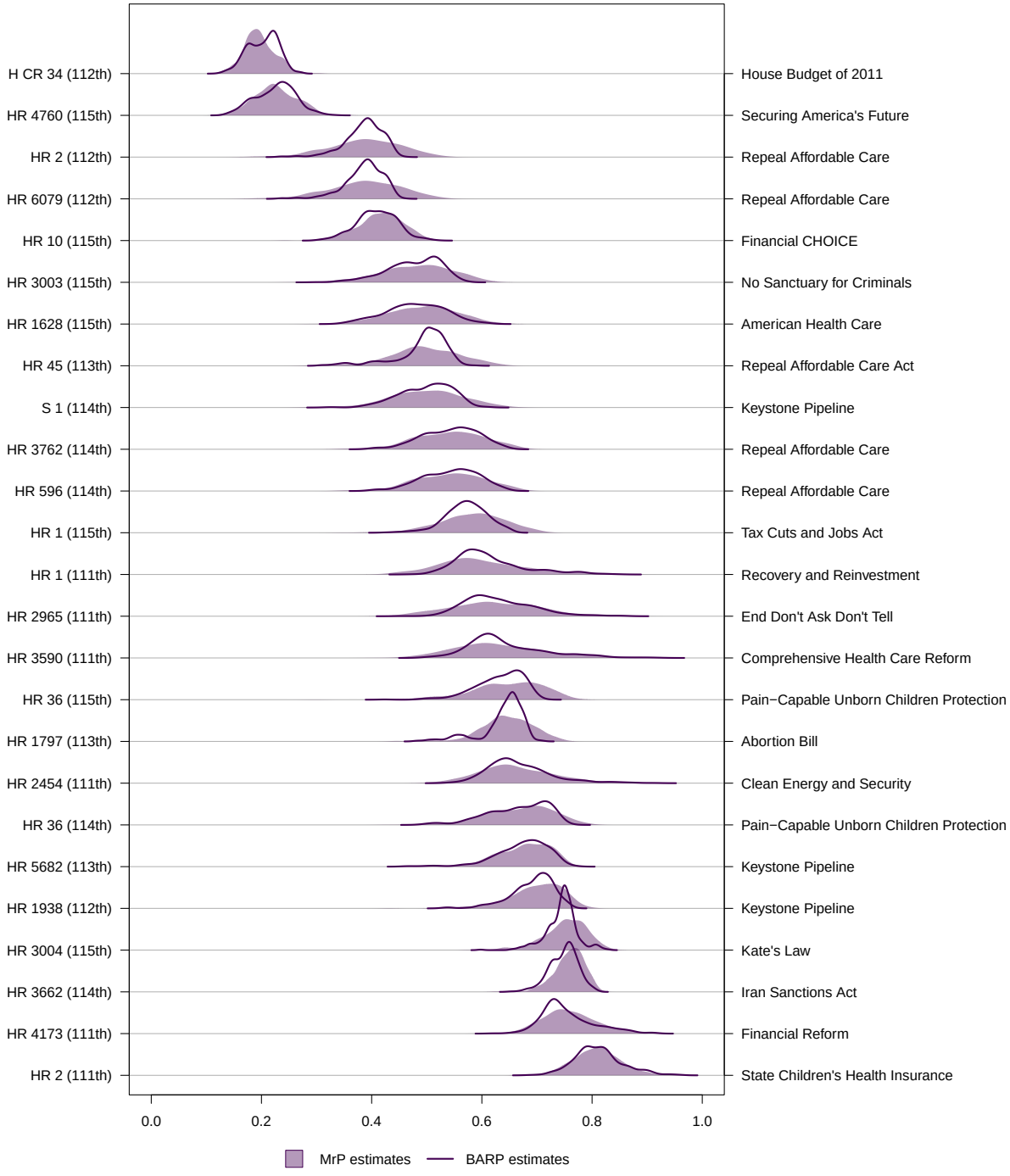


Figure A3: Distributions of district median voter preference estimates

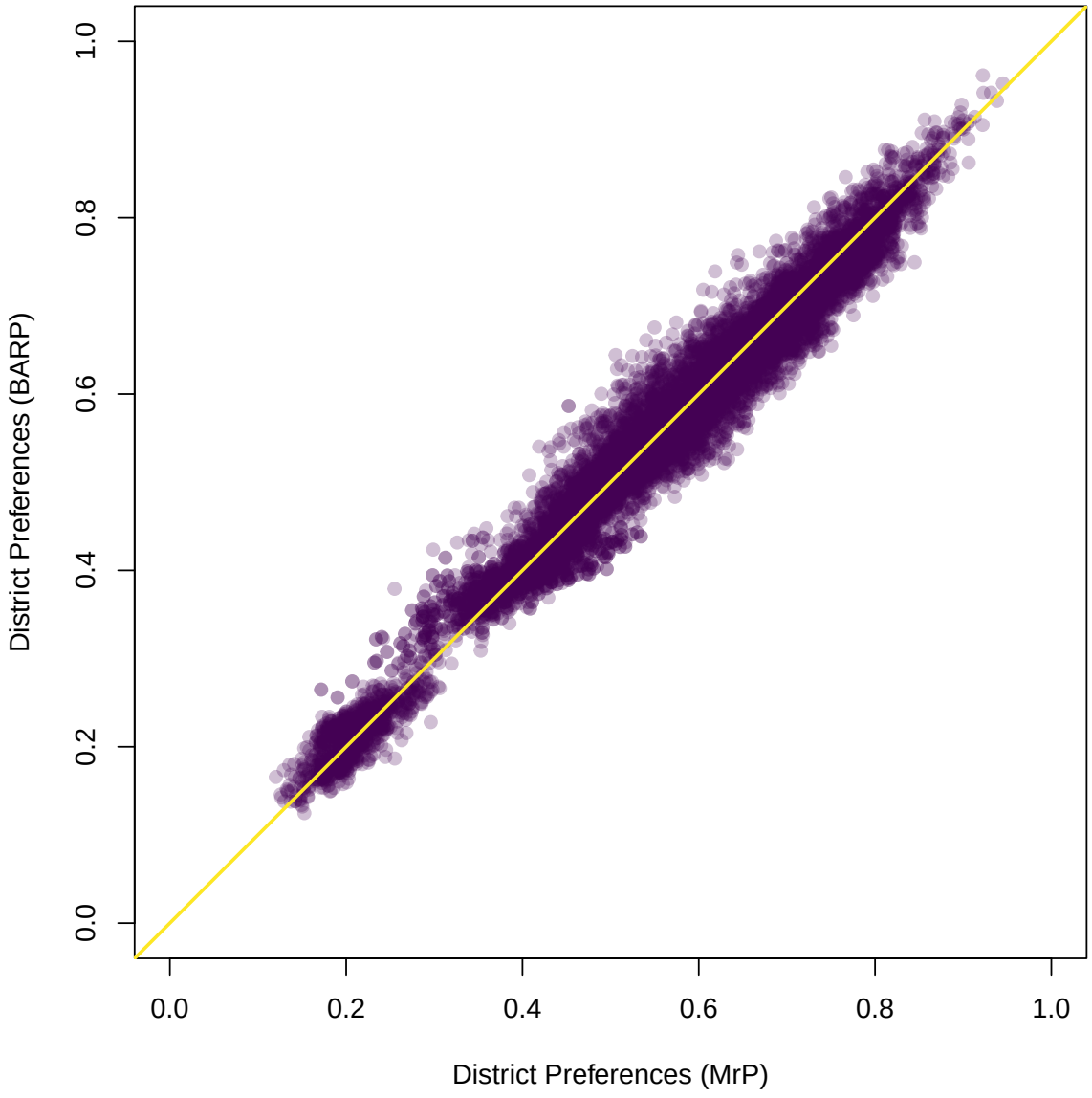


Figure A4: Scatterplot of MrP and BARP district preference estimates.

H Robustness

H.1 Alternative Time-Frames for Emphasis Aggregation

In our main analysis, we calculate emphasis scores based on the 30-second segment of a speech with the highest average emphasis score (see "Applying the Model to Footage of Key Vote Debates"). In this section, we extend our analysis by presenting results from additional models using emphasis scores derived from 10-, 20-, 30-, 40-, 50-, and 60-second segments, as well as the overall average emphasis. We replicate the Within-Between Random Effect Models (Models 1 and 2 from Table 4) using these alternative measures.

Figure A5 presents within- and between-effect coefficient estimates for the relationship between vote alignment and speech emphasis across different emphasis calculations, demonstrating that the results remain robust regardless of the emphasis aggregation procedure.

H.2 Binary Measure of District-Level Support

In this section, we employ binary measures of district-level support and district-level opposition to reevaluate the notion that legislators deliver more emphatic speeches when their vote is aligned with their electorate. Modeling district support as binary rather than continuous reflects the idea that it might be sufficient for legislators to know that their position is shared by the majority of their voters. Once legislators are confident that the majority of their electorate is on their side, it may be less important how large that majority is. Similarly, it may be sufficient for legislators to have a sense that a majority of their voters disagrees with them to keep them from delivering an emphatic speech on the floor.

The variable Aligned Vote distinguishes between legislators with high and low levels of alignment between their vote and the preferences of their district. The variable takes

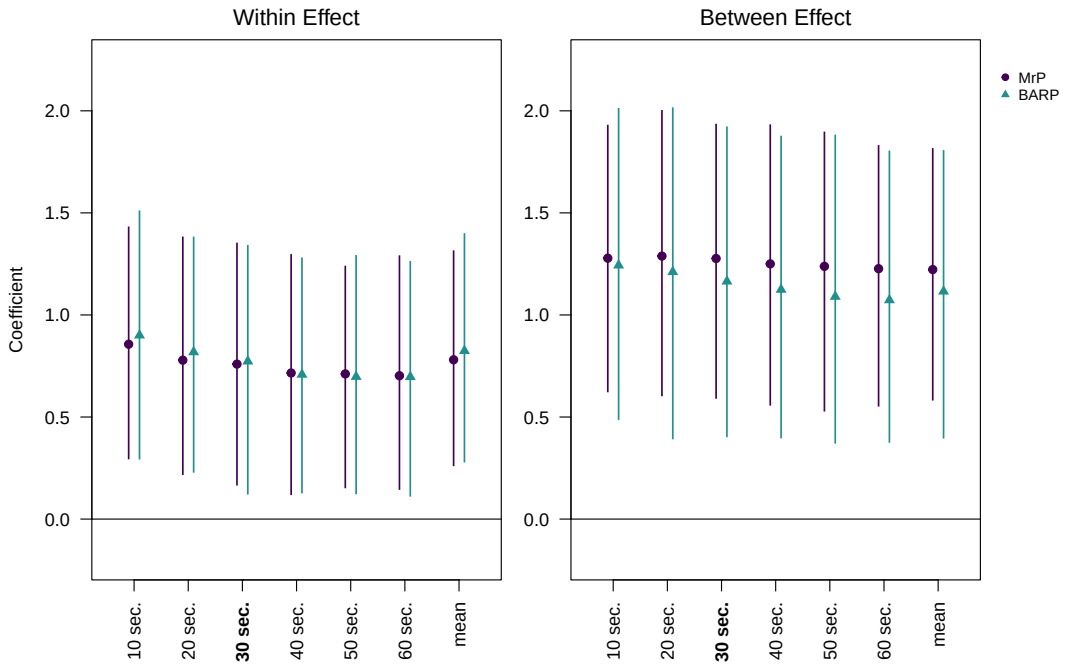


Figure A5: Point estimates and 95% wild bootstrapped confidence intervals of between effects of vote-alignment based on different computations of the dependent variable.

the value one if the estimated vote-district alignment is larger than a specific threshold t and zero if the estimate is lower than 0.45:

$$\textsc{Aligned Vote} = \begin{cases} 1 & \text{if } \textsc{District Alignment} > t,\ t \in [0.5, 0.7] \\ 0 & \text{if } \textsc{District Alignment} < 0.45 \end{cases} \tag{9}$$

The threshold t determines the level of alignment between a legislator and their district that is necessary to qualify as district support. Because the choice of this threshold is arbitrary, we vary it between 0.5 and 0.7 and estimate all models based on the different threshold values. Values of district alignment that fall between the two thresholds are treated as missing and the respective speeches are excluded from the analysis.

The variable Unaligned Vote is the inverse of district support and differentiates between legislators with low levels of support and those with high levels of support. It takes the value one if the district preference estimate is lower than a specific threshold t

and the value zero if the estimate is larger than 0.55. Again, the threshold value is varied between 0.3 and 0.5.

$$\textsc{Unaligned Vote} = \begin{cases} 1 & \text{if } \textsc{District Alignment} < t,\ t \in [0.3, 0.5] \\ 0 & \text{if } \textsc{District Alignment} < 0.55 \end{cases} \tag{10}$$

We employ two models to estimate the effect of district support and district opposition on the emphasis in legislative speech. First, we estimate linear models using ordinary least squares (OLS). District support and district opposition at varying levels of t enter the model as the independent variable of interest. We include the same set of control variables as before and estimate separate models for legislators who rise in support and in opposition to a bill. Second, to account for debate-level and legislator-level heterogeneity, we estimate multilevel models with varying intercepts for debates and legislators. We expect district support to have a positive effect on speech emphasis. District opposition is expected to negatively affect speech emphasis.

The results confirm the finding that legislators who rise in opposition to a bill deliver more emphatic speeches when the majority of voters in their district supports their vote. Figure A6 depicts the effect of district support and district opposition across different thresholds t from the OLS and multilevel models for legislators who rise in opposition to a bill. In line with expectations, the models show positive effects for district support and negative effects for district opposition. Based on the OLS model, if at least 55% of the district support the vote of a legislator, representatives are expected to deliver a speech that scores 0.30 [0.12, 0.46] points higher compared to a legislator with district support of less than 45%. Figure A6 shows that this effect is evident in both the OLS and in the multilevel specification and is stable across different values of t. Conversely, when legislators deliver a speech when the majority of voters in their district disagrees with their vote, they deliver less emphatic speeches than legislators who have the majority of their constituents on their side.

Turning to the speech delivery of legislators who rise in support of a bill, we find no

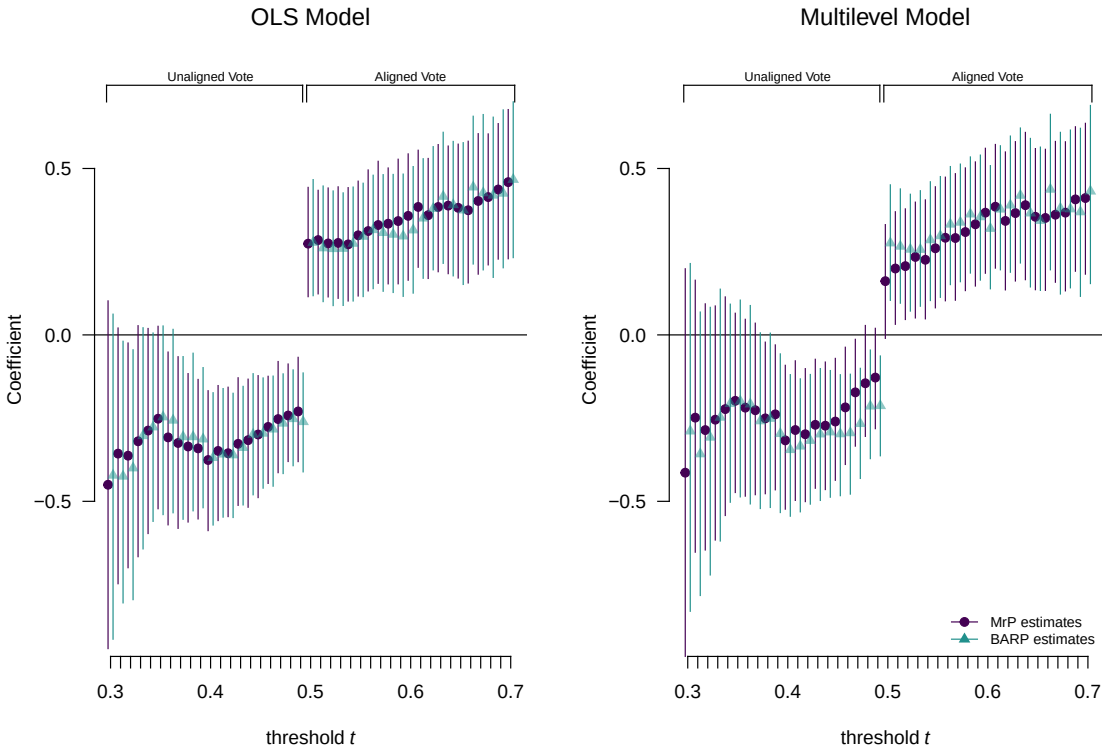


Figure A6: OLS and multilevel model coefficients and 95% wild bootstrapped confidence intervals for the effect of district opposition and district support on the emphasis in speeches in opposition to a bill.

evidence that they vary their delivery depending on public support for the bill. Figure A7 shows the effect of district support and district opposition across varying threshold values t from the OLS and multilevel models for legislators who rise in support of the bill. None of the effects is distinguishable from zero at conventional levels of statistical significance.

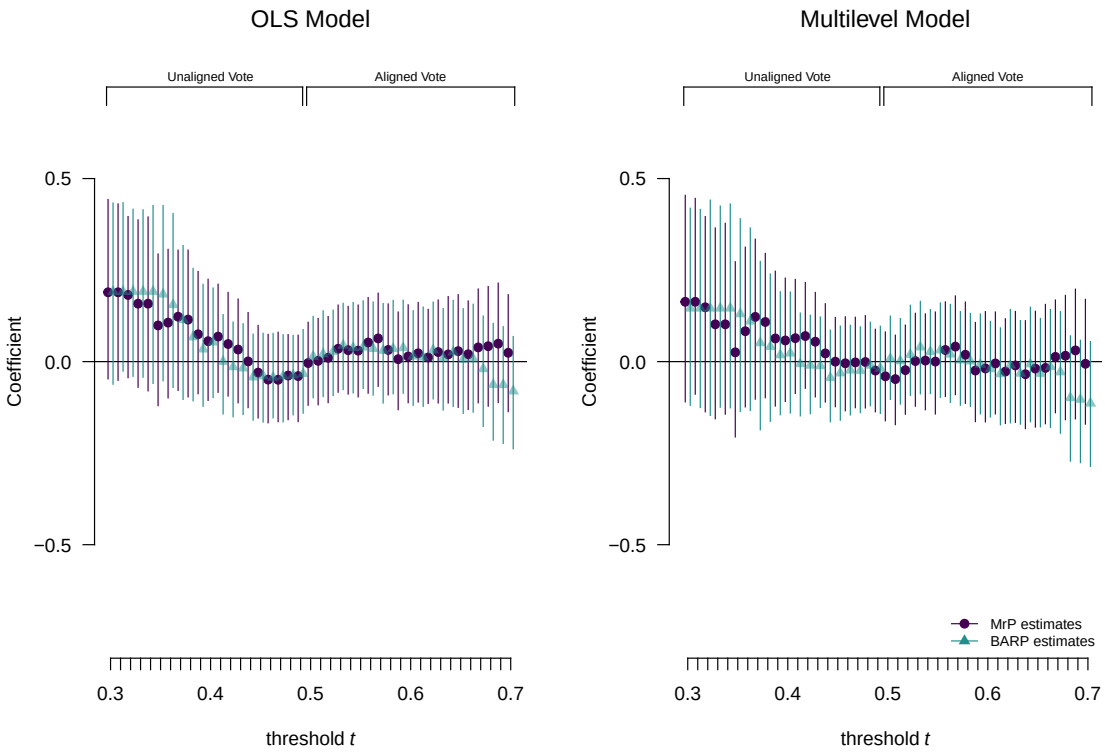


Figure A7: OLS and multilevel model coefficients and 95% wild bootstrapped confidence intervals for the effect of district opposition and district support on the emphasis of speeches in support of a bill.

I List of Data Sources

Data for District-Level Public Opinion Estimation

- Census Data
  - U.S. Census Bureau (2010)
  - U.S. Census Bureau (2012)
  - U.S. Census Bureau (2013*a*)
  - U.S. Census Bureau (2014*a*)
  - U.S. Census Bureau (2015)
  - U.S. Census Bureau (2016*a*)
  - U.S. Census Bureau (2017)
- Survey Data
  - CCES 2010 Common Content: Ansolabehere (2012)
  - CCES 2012 Common Content: Ansolabehere and Schaffner (2013)
  - CCES 2013 Common Content: Ansolabehere and Schaffner (2019)
  - CCES 2015 Common Content: Ansolabehere and Schaffner (2017*b*)
  - CCES 2016 Common Content: Ansolabehere and Schaffner (2017*a*)
  - CCES 2017 Common Content: Schaffner and Ansolabhere (2019)
- Congressional District Population Estimates: Manson et al. (2020)
- Congressional District Shapefiles: Lewis et al. (2013), U.S. Census Bureau (2013*b*), U.S. Census Bureau (2014*b*), U.S. Census Bureau (2016*b*), U.S. Census Bureau (2018).
- Religious Congregations and Membership Study, 2000: Ulmer (2019)
- Presidential Vote Share data: MIT Election Data and Science Lab (2017*b*)

Other

- Roll Call Votes: Retrieved from the official website of the Clerk of the United States House of Representatives (*Office of the Clerk, U.S. House of Representatives*, 2020).
- US Congress Legislators Data Set: GovTrack.us (2020)
- Voteview NOMINATE Ideology scores: Lewis et al. (2020)
- U.S. House of Representatives Election Results: MIT Election Data and Science Lab (2017*a*)

References

Ansolabehere, Stephen. 2012. "CCES Common Content, 2010.".
URL: *https://doi.org/10.7910/DVN/VKKRWA*

Ansolabehere, Stephen and Brian F. Schaffner. 2017*a*. "CCES Common Content, 2016.".
URL: *https://doi.org/10.7910/DVN/GDF6Z0*

Ansolabehere, Stephen and Brian Schaffner. 2013. "CCES Common Content, 2012.".
URL: *https://doi.org/10.7910/DVN/HQEVPK*

Ansolabehere, Stephen and Brian Schaffner. 2017*b*. "CCES, Common Content, 2015.".
URL: *https://doi.org/10.7910/DVN/SWMWX8*

Ansolabehere, Stephen and Brian Schaffner. 2019. "CCES Common Content, 2013.".
URL: *https://doi.org/10.7910/DVN/KPP85M*

Bisbee, James. 2019. "BARP: Improving Mister P Using Bayesian Additive Regression Trees." *American Political Science Review* 113(4):1060–1065.

Gelman, Andrew and Thomas C. Little. 1997. "Poststratification into Many Categories Using Hierarchical Logistic Regression." *Survey Methodology* 23(2):127–135.

GovTrack.us. 2020. "US Congress Legislators.". Accessed: 2020-10-22.
URL: *https://data.world/govtrack/us-congress-legislators*

Huang, Xuedong, Alex Acero and Hsaio-Wuen Hon. 2001. *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*. Prentice Hall.

Jones, Dale E., Sherri Doty, Clifford Grammich, James E. Horsch, Richard Houseal, mac Lynn, John P. Marcum, Kenneth M. Sanchagrin and Richard H. Taylor. 2002. *Religious Congregations and Membership in the United States 2000: An Enumeration by Region, State and County Based on Data Reported for 149 Religious Bodies*. Glenmary Research Center.

Lewis, Jeffrey B., Brandon DeVine, Lincoln Pitcher and Kenneth C. Martis. 2013. "Digital Boundary Definitions of United States Congressional Districts, 1789-2012.".
URL: *http://cdmaps.polisci.ucla.edu*

Lewis, Jeffrey B., Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin and Luke Sonnet. 2020. "Voteview: Congressional Roll-Call Votes Database (2020).".
URL: *https://voteview. com*

Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler and Steven Ruggles. 2020. "IPUMS National Historical Geographic Information System: Version 15.0.".

MIT Election Data and Science Lab. 2017*a*. "U.S. House 1976–2018.".
URL: *https://doi.org/10.7910/DVN/IG0UN2*

MIT Election Data and Science Lab. 2017*b*. "U.S. President 1976–2020.".
URL: *https://doi.org/10.7910/DVN/42MVDX*

Office of the Clerk, U.S. House of Representatives. 2020. Accessed: 2020-10-22.
URL: *https://clerk.house.gov*

Schaffner, Brian and Stephen Ansolabhere. 2019. "CCES Common Content, 2017.". **URL:** *https://doi.org/10.7910/DVN/3STEZY*

Taylor, Luke and Geoff Nitschke. 2018. Improving Deep Learning with Generic Data Augmentation. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*. pp. 1542–1547.

Ulmer, Gail. 2019. "Religious Congregations and Membership Study, 2000 (State File).". **URL:** *https://doi.org/10.17605/OSF.IO/Q8EMK*

U.S. Census Bureau. 2010. "2010 American Community Survey.".

U.S. Census Bureau. 2012. "2012 American Community Survey.".

U.S. Census Bureau. 2013*a*. "2013 American Community Survey.".

U.S. Census Bureau. 2013*b*. "Cartographic Boundary Files - Shapefile, Congressional Districts: 113th Congress.". **URL:** *https://www.census.gov/geographies/mapping-files.html*

U.S. Census Bureau. 2014*a*. "2014 American Community Survey.".

U.S. Census Bureau. 2014*b*. "Cartographic Boundary Files - Shapefile, Congressional Districts: 114th Congress.". **URL:** *https://www.census.gov/geographies/mapping-files.html*

U.S. Census Bureau. 2015. "2015 American Community Survey.".

U.S. Census Bureau. 2016*a*. "2016 American Community Survey.".

U.S. Census Bureau. 2016*b*. "Cartographic Boundary Files - Shapefile, Congressional Districts: 115th Congress.". **URL:** *https://www.census.gov/geographies/mapping-files.html*

U.S. Census Bureau. 2017. "2017 American Community Survey.".

U.S. Census Bureau. 2018. "Cartographic Boundary Files - Shapefile, Congressional Districts: 116th Congress.". **URL:** *https://www.census.gov/geographies/mapping-files.html*

Warshaw, Christopher and Jonathan Rodden. 2012. "How Should We Measure District-Level Public Opinion on Individual Issues?" *The Journal of Politics* 74(1):203–219.