Pay attention to this! Explaining emphasis in legislative speech using automated video analysis in the US House of Representatives

Oliver Rittmann* Tobias Ringwald† Dominic Nyhuis‡

Last updated:
December 7, 2020

Abstract

Why do legislators sometimes deliver passionate speeches and sometimes tedious monologues? We argue that legislators make passionate appeals when they want to signal support or opposition to a bill. Whether legislators choose to send such a signal depends on the preference of the median voter in their districts. We expect legislators to deliver more emphatic speeches if their floor vote is aligned with the preferences of their electorate. To test this argument, we apply automated video analysis to plenary recordings of speeches on key votes in the 111th–115th US House of Representatives (2009–2018). We match the speech emphasis with district preferences on the 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 choice. The paper discusses the implications of these findings for our understanding of legislative debates.

*School of Social Sciences, University of Mannheim. E-mail: `orittman@mail.uni-mannheim.de`. Research for this project was supported by the *Deutsche Forschungsgemeinschaft* and the *Sonderforschungsbereich 884*. Further support is gratefu lly acknowledged from the research alliance *ForDigital* with funding from the *Land Baden-Württemberg*. We thank Morten Harmening, Felix Münchow, and Marie-Lou Sohnius for excellent research assistance. We are grateful to Thomas Gschwend, Lukas Stoetzer, Christian Arnold, Seo-young Silvia Kim, Patrick Kraft, Marcel Neunhoeffer, Anna Aadendorf, Oke Bahnsen, Sean Carey, Franziska Quoß, Viktoriia Semenova, Christoph Steinert, and participants of the Comparative Politics Seminar at the University of Washington for feedback on earlier versions of this manuscript.

†Computer Vision for Human-Computer Interaction Lab, Karlsruhe Institute of Technology.

‡Department of Political Science, University of North Carolina at Chapel Hill.

Introduction

Research on political speech has made tremendous progress in recent years. The widespread availability of digitized political text has allowed researchers to gain important insights into the use and substance of political speech and legislative speech in particular (Diermeier et al., 2011; Lauderdale and Herzog, 2016; Monroe et al., 2008; Proksch and Slapin, 2012, 2014). Yet, despite the value of these efforts for legislative research, key dimensions of political speech are systematically disregarded in this literature. Given their focus on the textual features of political speech, the existing studies have paid little attention to the nonverbal characteristics of legislative speech. While focusing on the substance of political speech may well be sufficient for studying important political science questions on the strategic position-taking of legislators, it ignores key functions of political speech. By relying on oratory tactics, legislators try to sway their colleagues and appeal to the general public. While some of these tactics may shine through in the written word, most of the nonverbal appeal is lost in transcription. Therefore, in an effort to help fill this gap in the research on political speech, we focus on the nonverbal characteristics of legislative speech.

Starting from the empirical observation that legislators sometimes deliver tedious monologues, while they give rousing speeches at other times, this paper asks what explains such variation in the delivery of legislative speech. To answer this question, we build on work on the responsiveness of legislative speech to public opinion and on the use of legislative speech for the purpose of signaling (Bäck and Debus, 2018; Baumann et al., 2015; Hill and Hurley, 2002; Umit and Auel, 2020). We argue that legislators are not only mindful of public opinion in what they say, but that they are also strategic in how they say it. As most legislative speeches go all but unnoticed by the public, legislators occasionally make emphatic appeals to increase the odds of being featured in the media. Particularly in the current media environment, it is imperative to deliver a good soundbite in order to make it past the media gatekeepers or to go viral on social media (Esser, 2008; Esser and Strömbäck, 2014; Larsson, 2020; Negrine and Lilleker, 2002; Strömbäck, 2008). Legislators are aware of this bottleneck and they use it to their advantage by strategically delivering fiery speeches when they want to signal their position to their constituents. Therefore, above and beyond the substantive signals that legislators send in their speeches, their delivery sends important signals in their own right. Whether legislators choose to send such a signal depends on the policy preferences of their constituents. We argue that legislators will only highlight their position using emphatic speech when their constituents hold a strong preference for or against a particular proposal and when that preference is aligned with the preference of the legislator.

To trace this proposition empirically, we rely on recent methodological innovations for the study of audio and video data. Using a convolutional neural network, we analyze video footage from the US House of Representatives to systematically gauge the nonverbal aspects of legislative speech in a large-n study. To study the effect of public preferences on the delivery of speeches, we use Multilevel Regression and Poststratification (MrP) and Bayesian Additive Regression Trees and Poststratification (BARP) for estimating district preferences (Warshaw and Rodden, 2012; Bisbee, 2019). Specifically, we estimate district preferences on a series of congressional bills from the 111th to the 115th US House of Representatives (2009–2018) using data from the Cooperative Congressional Election Study (CCES). The resulting estimates are matched with video footage from speeches during plenary debate on these bills. The results support the idea that legislators employ emphatic appeals to signal their position to their constituents under favorable conditions.

The results have important implications for our understanding of legislative speech. Our study is one of few contributions that take the nonverbal characteristics of legislative speech serious. In addition to showing that the nonverbal characteristics contain useful information for political science research, we highlight how nonverbal features of political speech are shaped by strategic considerations. These findings may prove particularly valuable for researchers trying to gauge the substance of political conflict from political speech (Bäck and Debus, 2018; Diermeier et al., 2011; Lauderdale and Herzog, 2016; Monroe et al., 2008; Proksch and Slapin, 2009). Incorporating the nonverbal characteristics into these efforts can generate new insights as emphatic legislative speech helps distinguish key policy statements from everyday speech.

Methodologically, the paper speaks to a developing research in political science which has adopted methodological innovations to extract nontextual information from audio and video recordings (Dietrich et al., 2019b; Knox and Lucas, forthcoming). While previous contributions in this field have focused on descriptive relationships and the face validity of the new measures, our study is among the first to present a theoretical mechanism to explain variance in the nonverbal characteristics of political speech.

Moving beyond the textual features of political speech

There is a growing interest in the analysis of political speech. Particularly the comprehensive digitization of parliamentary records has helped expand our understanding of the use (Maltzman and Sigelman, 1996; Morris, 2001; Proksch and Slapin, 2012) and substance (Hill and Hurley, 2002; Morris, 2001; Quinn et al., 2010) of parliamentary speech. Despite the importance of this research program for legislative politics, it is subject to notable limitations. Efforts to categorize political speech have almost exclusively relied on their textual features. While focusing on the text and hence on the substance of speeches is a reasonable choice for many research questions, legislative speech has important dimensions that are difficult to study on the basis of 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 our understanding of their intent and their effects. Succinctly put, research has learned a lot about *what* legislators say, but not *how* they say it.

While previous research has focused on the substance of legislative speech, some contributions have attempted to quantify the non-textual aspects of legislative speech (Banning and Coleman, 2009; Bucy, 2016; Wasike, 2019) and how they affect perceptions of the speaker (Burgoon et al., 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 area is to build on the recent advances for the automated analysis of audio and video data and to apply these methodological innovations to the ever more widely available digitized recordings of legislative speech.

A nascent literature has begun to employ these tools to research the nonverbal characteristics of legislative speech and other recordings of political interest. Analyzing vocal pitch in audio recordings from the US House of Representatives, Dietrich et al. (2019b) demonstrate that female legislators speak with greater intensity about women. Dietrich and Juelich (2018) also rely on vocal pitch to show that candidates in a televised debate exhibited higher pitch when speaking about issues owned by their respective parties. Dietrich et al. (2019a) even provide evidence that the vocal pitch of Supreme Court Justices during oral argument is predictive of their vote choices. Audio recordings of the Supreme Court are also analyzed in the work by Knox and Lucas (forthcoming), who introduce a general model of audio data and apply it to judicial speech in an attempt to classify whether Justices express skepticism during oral argument.

Political science applications using digitized video recordings are even rarer than the few studies using audio data. Dietrich (2015, forthcoming) 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. Joo et al. (2019) highlight the opportunities for political communication scholars to automatically classify nonverbal behavior from digitized video data by studying footage from a televised candidate debate.

While these studies constitute valuable efforts in moving 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 the tools used 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 variance in the nonverbal aspects of legislative speech. To this end, we develop and test a theoretical account for emphatic legislative speech.

Emphatic legislative speech as signaling

To explain the variation in the nonverbal aspects of political speech, we begin by defining the concept of interest: the *emphasis in legislative speech*. While emphasis is a multifaceted property of political speech, there are key elements that characterize emphatic legislative speech. The most important nonverbal features of emphatic political speech are extensive gesturing, high volume and a fast-paced delivery.

In order to construct a theoretical account for the varying emphasis in legislative speech, we build on the strategic behavior of legislators. One of the most well-established patterns in legislative politics is the link between re-election concerns and legislator behavior—the electoral connection (Mayhew, 1974). Beyond behaviors with immediate policy consequences such as roll call voting or pork barreling, legislators have been shown to be mindful of public opinion in their communication (Grimmer, 2013a,b; Grimmer et al., 2015) and in legislative speech specifically (Hall, 1996; Highton and Rocca, 2005; Hill and Hurley, 2002; Morris, 2001). These studies highlight that legislators use speeches to signal that their policy positions are aligned with the preferences of their constituents. Compared with other legislative instruments, speeches allow legislators to signal that they not only hold the right positions, but that they also care about the right issues (Kalaf-Hughes, 2020).

Yet, even though there is consistent evidence that legislative speech is responsive to public opinion, research has paid much less attention to the question which strategies legislators employ to be heard by the public. While legislators clearly use speeches and other legislative activities to signal to the public, the majority of these signals go all but unnoticed. Although legislative speech is a matter of public record, the likelihood that the public seeks out these signals is negligible. Therefore, the only way for signals in legislative speech to reach the public is if they are amplified by traditional or social media. To increase the odds of being featured in traditional media or going viral on social media, legislators try to create good soundbites. By making passionate appeals on the plenary floor, legislators seek to overcome the lack of public attention. The upshot of extending the strategic considerations from the substance of legislative speech to the speech delivery is that legislators make fiery appeals in their speeches under specific and predictable circumstances in the hopes of becoming visible to the public, while they will opt for a monotone delivery when they try to fly under the radar.

The proposed link between legislator efforts and media coverage ties in well with a broader research agenda on the visibility of legislators in the media. Whereas research on the ability of legislators to receive media coverage has traditionally focused on the effects of formal roles (Cook, 1986; Squire, 1988), more recent research on the subject has identified a variety of factors that shape the ability of legislators to attract media attention (Vos, 2014). Not only do legislators pursue a variety of strategies to be featured in the media (Gershon, 2012; Lipinski and Neddenriep, 2004; Sellers and Schaffner, 2007), there is strong evidence that legislators' communication skills affect their success in receiving media attention (Amsalem et al., 2017, forthcoming; Sheafer, 2001, 2008; Sheafer and Wolfsfeld, 2004; Wolfsfeld and Sheafer, 2006). These results are closely mirrored in the social media realm where emotional and personal appeals are associated with the success and virality of messages (Heiss et al., 2019; Nave et al., 2018). These contributions clearly show that legislators are strategic in their efforts to receive media coverage and that emphatic appeals are a promising strategy for receiving media attention. Importantly, our interest is whether legislators choose a particular delivery style in an effort to signal their position to the public, not whether such efforts are ultimately successful.

Building on the research on the electoral connection in the substance of legislative speech, we study the effect of constituency opinion on the efforts of legislators to signal their position to the public. In line with an economic voting account (Downs, 1957), legislators' utility of signaling their position depends on the preferences of their electorate towards the bill under consideration. As voters value policy positions that approximate their own preference, a policy signal is beneficial to legislators when that signal is in line with the preference of the median voter, leading us to expect emphatic appeals in cases of alignment.

Distinguishing between policy signals in legislative speech and the underlying vote choice opens up important nuance in the analysis of strategic legislative behavior. For a variety of reasons, legislators may choose to cast a vote that is not aligned with the preferences of their district. For instance, legislators may hold strong preferences on a proposed bill that goes against the preferences of their districts. Alternatively, legislators might face partisan pressures which make a vote against the party line costly (Nokken, 2000; Grose and Middlemass, 2010), particularly in the current climate of elite polarization. While voting against the district is a potential source of vulnerability for legislators, there is a good chance that such a vote might go unnoticed. Yet, while a vote against the district may well fly under the radar, legislators should not go out of their way to signal such a vote to their constituents.

By the same token, distinguishing between the substance of legislative speech and its delivery allows a more nuanced account of legislator strategy. In the same example as before, legislators may cast a vote against the preferences of their constituents. This may result in legislators wanting to take the floor to explain and defend their vote choice in case an interested party becomes aware of it. At the same time, legislators have little incentive to deliver an impassioned appeal as this constitutes the danger of unwanted public attention for an unpopular decision.

Consequently, the delivery of legislative speech contains valuable cues about legislator strategy above and beyond the underlying vote choice and the substance of legislative speech. Using emphatic appeals as signaling devices, legislators are expected to send policy signals if their constituents have a clear preference for or against a bill and if that preference matches the vote choice of the legislator.

Research design

Are legislators more likely to deliver emphatic legislative speeches if the preferences of their districts on a bill align with their roll call vote? To test this proposition, we study debates on 25 pieces of legislation in the 111th–115th US House of Representatives (2009–2018). This sample was selected with four criteria in mind. The bill must be of high salience, survey data to assess public opinion towards the bills must be available, there should be some partisan conflict on the bill, and public opinion towards the bill must vary across 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 US House of Representatives. These bills overlap to a large extent with votes that were classified as "key votes" by *Congressional Quarterly*. As such, they are of high salience and cover a wide range of domestic and foreign policy issues (Ansolabehere and Jones, 2010). From this sample we discard bills that were passed with no partisan conflict[1] and bills without variation of district-level public opinion on the bill. We do this as we consider partisan conflict as a necessary condition for our argument: If there are no diverging positions either among legislators or among the electorate, then signals in legislative speech lose their value. Table 2 lists the 25 bills in the resulting sample, brief summaries of their contents are provided in the Appendix. On a theoretical level, the restriction to salient and partisan votes is

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

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 even more unlikely to be observed by the public.

In the remainder of this section, we first introduce the dependent variable, the emphasis in legislative speech, and how it can be gauged from video footage using computer vision. Next, we discuss the estimation of district preferences on key pieces of legislation as the independent variable. We close this section with a discussion of the statistical model for estimating the effect of district preferences on legislative speech.

Measuring emphasis in legislative speech using automated video analysis

To study the emphasis in legislative speech, we analyze video recordings of key vote debates in the US House of Representatives. We compile video recordings of these debates from *HouseLive*.[2] The sample contains video recordings of all debates on the 25 pieces of legislation in our study. We manually discard irrelevant sequences to ensure that we only analyze footage where the camera fully captures legislators who are delivering a speech.[3] This results in 78 hours of video footage comprising 2,383 speeches by 548 legislators. Table 5 in the Appendix lists the respective debates and pieces of legislation.

We employ computer vision to measure emphasis in political speech from video recordings. Specifically, we generalize manual annotations based on a set of training videos to all videos in the sample. We start by drawing an additional and independent sample of 245 speeches by 116 legislators on 37 bills from the 115th Congress as training/test footage. The videos were split into 184 training and 61 test videos. Four trained student assistants were tasked with annotating the emphasis in the speeches using a 7-point Likert scale, ranging from -3 (very low) to $+3$ (very high), for every non-overlapping two-second segment.[4] To code the videos, raters accessed a website, where they could play the training/test videos. Raters annotated the emphasis of the speeches as the videos were playing in a set-up comparable to real-time response measurement (Maier et al., 2006, 2016).

Every video was annotated by two randomly selected coders in order to better judge the emphasis in the video and to evaluate inter-rater agreement. As continuous annotation of

[2] *HouseLive* is an online service that provides archived video footage of proceedings in the US House of Representatives dating back to January 2009. houselive.gov was recently taken offline and replaced with live.house.gov, which only provides video recordings dating back to spring 2017 at the time of writing. However, the former website is still publicly available in the web archives of the Library of Congress at www.loc.gov.

[3] Detailed video cutting rules are documented in the Appendix.

[4] The coding scheme for the manual annotations is documented in Table 4 in the Appendix.

video data is subject to reaction time and mental processing speed, annotations can move out of sync. Therefore, we align the annotation sequences by the two coders using the mean absolute error distance. The alignment process shifts values by at most two seconds, i.e. by one segment. Table 1 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.

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 speech emphasis scores for each two-second segment of the legislative speeches. As context information is useful for predicting the current emphasis state of a speaker, we include the surrounding two-second segments for the prediction task. 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. Regarding the video input, this includes resizing, normalizing, and randomly cropping the input images. Random cropping helps to 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).[5] The final video input are images with a size of 299×299 pixels. Regarding the audio input, we extract 20 Mel-frequency cepstral coefficients (MFCCs) from the raw audio data and feed it into the network (cf. Huang et al., 2001, ch. 6).

To predict the emphasis in the speeches, we employ a convolutional neural network (CNN). 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 leverages the features learned in the first stage to predict 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 in computer vision. Therefore, especially in cases with small training data sets, pre-trained networks are commonly used for feature extraction and have proven to be highly effective (Chollet and Allaire, 2018; Carreira and Zisserman, 2017).

[5] To ensure reproducibility, we use center cropping for the test footage.

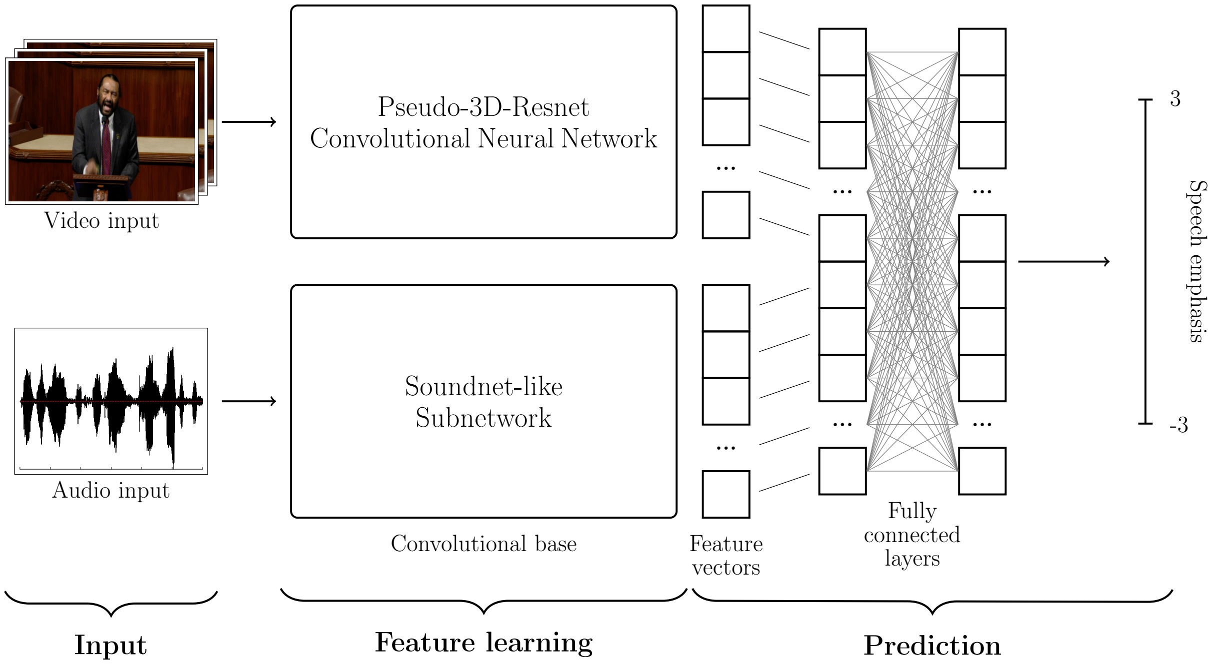

Figure 1: Convolutional neural network architecture

As the scope of our training data is limited, we use pre-trained networks for feature extraction for both the video and the audio input. For the video input, we use a state-of-the art pseudeo-3D-Resnet CNN (Qiu et al., 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 the 299×299 pixels images as input and generates a 2048-dimensional feature vector. For the audio input, we use a Soundnet-like subnetwork (Aytar et al., 2016). This network takes the 20 MFCCs as input and generates a 512-dimensional audio-feature vector.

After passing both our 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 concatenate both vectors to a single vector and pass it 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 summarizes 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). Dropout is an effective and widely used technique to prevent neural networks from overfitting. Applying dropout essentially means randomly setting a number of output features of a layer in a neural network to zero during the training phase. The idea is to add noise to the output values to prevent the network from picking up patterns that are unique

Table 1: 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
Number of videos	184	61			
Number of annotated segments	12,686	3,720			
Mean absolute error	0.438	0.438	1.192 ± 0.013	0.932	0.552
Lin's concordance coefficient	0.816	0.816	-0.001 ± 0.016	0.000	0.764
Pearson's correlation coefficient	0.816	0.818	-0.001 ± 0.016	0.000	0.770

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

to the training data. We use the Adam algorithm to train the model (Kingma and Ba, 2014).

Model evaluation

We now turn to the evaluation of the neural network. To that end, we apply the trained model to the held out test set of 61 videos and compare the model predictions with the human annotations. Table 1 shows the results of this comparison, along with the results based on random guessing (drawing predictions from a clipped standard normal distribution) and zero guessing (always 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. Considerably lower MAE values are achieved with the neural network scores. Based on the MAE, the machine prediction is 0.552 scale points off from the human annotation. This figure is close to the human inter-rater MAE. Unlike the guessed values, the model predictions show a high correlation for both the CCC and the PCC metrics. As before, the correlation between the machine prediction and the human annotators lies in the same range as the correlation between the human annotators. We thus conclude that the neural network reliably predicts the speech emphasis.

Applying the model to footage of key vote debates

We apply the trained model to the video footage for all key vote debates in our sample. The network predicts emphasis scores for each two-second video 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 score for each speech. However, this approach would ignore that speeches differ considerably 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 seems sensible to focus on shorter sequences within speeches. Legislators are aware that only short excerpts from 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, to score the speeches, we select the 30-second sequence with the highest within-speech average emphasis for each video. Specifically, we calculate the average emphasis scores for all possible 30-second sequences in a speech and choose the highest value to represent the speech.[6] 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.

Table 2 provides summary statistics for the resulting data by legislative debate. The number of legislators who delivered a speech 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 about the overall distribution of the emphasis scores across all debates, which range from -1.5 to 2.2. Thus, the distribution does not reach the endpoints of the emphasis scale that runs from -3 to $+3$. This is not surprising as we average over 30-second segments. The distribution can be characterized as approximately normal with a mean of 0.25 and a standard deviation of 0.73.

Estimating district-level bill preferences

Our key independent variable is the extent to which legislators' vote choices align with the preference of the constituents in their districts. We draw on survey data from multiple waves of the Cooperative Congressional Election Study (CCES) to estimate district-level preferences on the pieces of legislation. Each survey contains multiple questions on specific bills.[7] Respondents are provided with the title and a short description of the bill and are

[6] As the cutoff of 30 seconds is somewhat 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 scores, which has no effect on the substantive conclusions.

[7] Hill and Huber (2019) have recently highlighted the challenge of estimating public opinion on specific bills with survey data. While the authors voice a valid concern, the resulting data is preferable over alternative measures of district preference as it can be more easily tied to specific pieces of legislation.

Table 2: Summary statistics on the bills

Bill	Term	Title	Speakers	House vote	Emphasis Mean	Emphasis 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 displays the result of the final vote on the bill in the House.
Mean provides the mean emphasis scores across all speeches on a bill,
SD displays the associated standard deviation.

asked how they would have voted.[8] Matching each bill in our sample to a respective CCES item enables us to estimate the public opinion towards the bills in legislators' electoral districts.[9]

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 the district level to estimate district-level preferences is likely to yield biased estimates. We pursue two approaches to overcome this problem. First, we rely on the widely used strategy to estimate district preferences with multilevel regression and poststratification (MrP) (Gelman and Little, 1997; Lax and Phillips, 2009; Warshaw and Rodden, 2012). The basic idea behind MrP is to model individual survey responses as a function of

[8]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."

[9]Column 2 in Table 5 in the Appendix reports the matches of the CCES items with the debates in our sample.

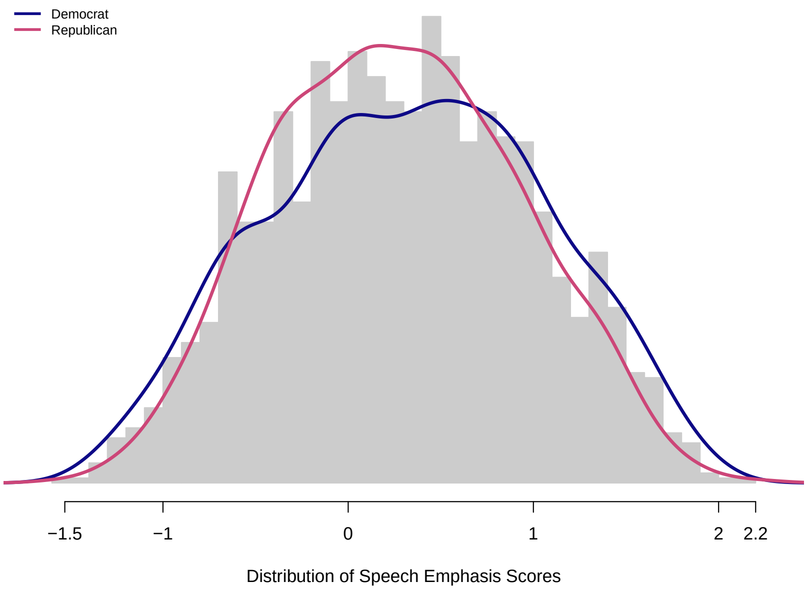

Figure 2: Histogram of emphasis scores.

Note: The endpoints of the x-axis denote the minimum and maximum values of the distribution.

demographic and geographic predictors in a multilevel model (Gelman and Hill, 2006) and to use the model to predict answers for respondents of all possible demographic-geographic combinations. To derive estimates for specific geographic units—in our case congressional districts—these predictions are then weighted by the percentage of respondent types in each congressional district. MrP has been shown to outperform simple disaggregation and provides reliable estimates for small-area public opinion (Warshaw and Rodden, 2012).[10]

Second, to strengthen the reliability of our analysis, we complement the estimates from the MrP approach with estimates from Bayesian regression trees and poststratification (Bisbee, 2019) (BARP). BARP relies on the same logic as MrP but replaces the multilevel model with a fully nonparametric regularization technique, Bayesian additive regression trees. In contrast to MrP, BARP allows for deep interactions between prognostic covariates even if they are not specified in the functional form of the model. This makes BARP less vulnerable to model misspecification and thus an optimal approach to validate the results of the analysis.

We employ MrP and BARP to estimate district-level preferences for each roll call in our sample. This results in a data set containing district-level preferences for 25 roll calls. The estimates fall between zero and one with higher values indicating high levels of support for a specific piece of legislation.

In the next step, we match this data to the representatives and their roll call record

[10]Details on the estimation of district preferences are provided in the Appendix.

on the 25 votes. Substantively, we are interested in the extent to which legislators' votes align with the preference of their electorate on the bills. To compute a suitable measure, we code roll call votes as one for legislators who voted in favor of a bill and zero for those who voted against it. As the district-level estimates range between zero and one, roll call records and district-level preferences now have a common scale. To assess the extent to which legislators' votes align with the preferences of their electorate, we calculate the absolute difference between a legislator's vote ($\textsc{Yes} = 1, \textsc{No} = 0$) and the preference of his or her district and substract this from the maximum distance, 1. We label this variable VOTE-DISTRICT ALIGNMENT:

$$\textsc{Vote-District Alignment} = 1 - |\textsc{Yae Vote} - \textsc{District Preference}| \tag{1}$$

Values close to 1 indicate high levels of alignment between a legislator and his or her district, values close to 0 indicate high levels of disagreement. On the extreme ends, for legislators who vote YES, the variable takes on the value 1 if all voters in their district support the respective bill. It takes the value 0 if a legislator votes YES while all voters in their district oppose the underlying piece of legislation and *vice versa*.

We present the distributions of the VOTE-DISTRICT ALIGNMENT variable from both estimation techniques by bill and legislators' vote choice in Figure 3. The bright density curves depict distributions of VOTE-DISTRICT ALIGNMENT by bill for legislators who voted yes, the dark densities depict the distributions of those legislators who voted no. We first note that for most bills MrP and BARP estimation result in very similar distributions of the alignment between legislator votes and district preferences. However, considerable differences in few instances underline the need for an examination of the sensitivity of the results to the underlying estimation techniques. Second, we note that for most bills the distributions of legislators who voted yes and no on a bill diverge. Consider for example the State Children's Health Insurance Act (HR 2, 111th) for which we estimate an average district-level support of 0.66 (SD = 0.08). Although there is variation across districts, almost all districts were rather supportive of the bill. Thus, VOTE-DISTRICT ALIGNMENT values of legislators who voted in favor of the bill are substantially higher than the values for legislators who voted against the bill. We return to this observation in the discussion of our statistical model.

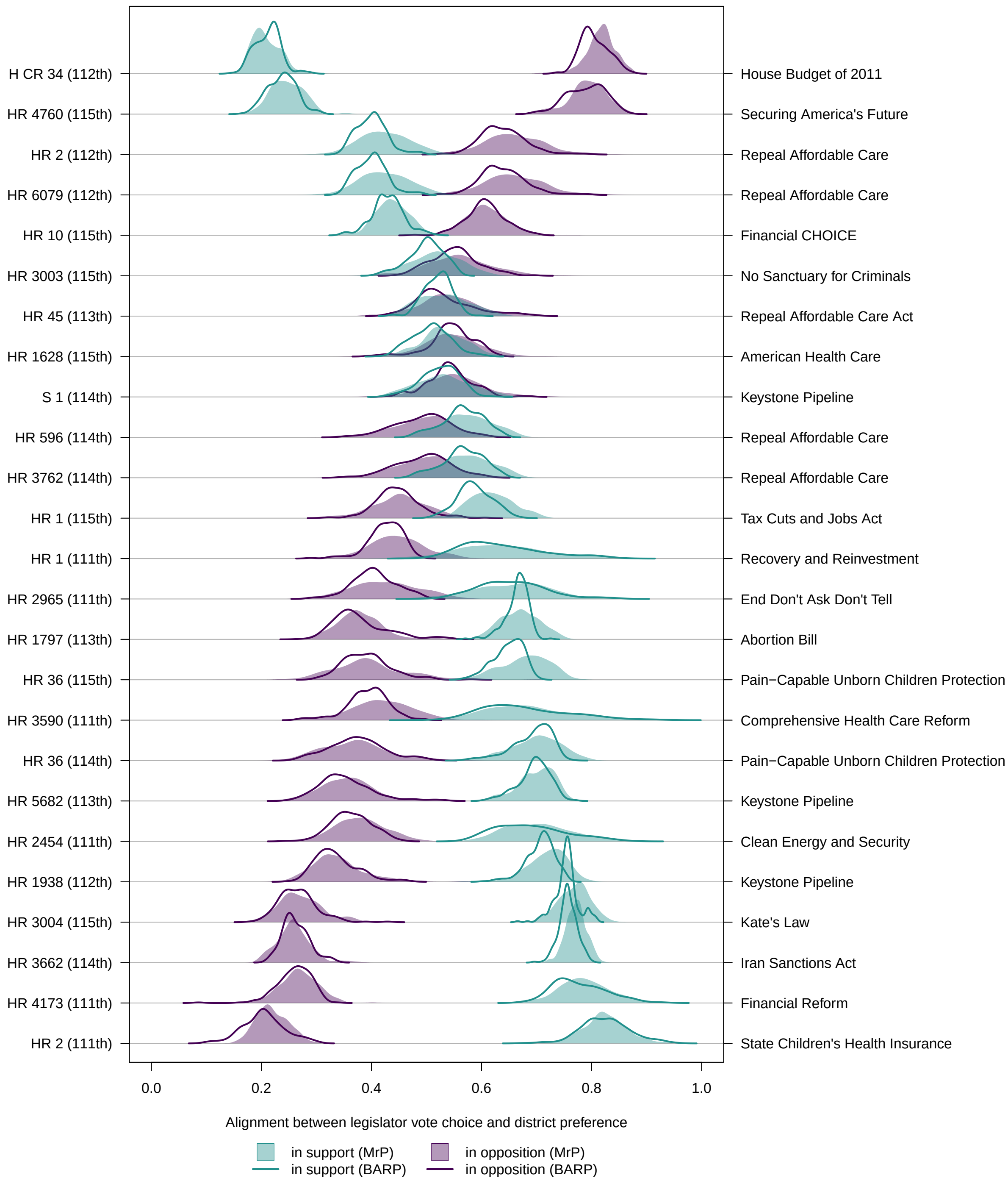

Figure 3: Distributions of alignment between legislators' roll call votes and the preferences of their electoral districts.

Modeling strategy

We now outline the statistical model which aims to identify the effect district preferences on legislators' emphasis in legislative speech. The unit of observation are legislative speeches. Speeches are clustered in two higher levels: individual legislators and legislative debates. To facilitate causal identification and to ensure that this clustering does not confound the estimates, the statistical model needs to meet two requirements. First, it needs to account for the heterogeneity of speech by different legislators. Second, it needs to account for systematic differences of public opinion between debates. The latter becomes apparent when looking at figure 3: There is considerable heterogeneity in the alignment between public opinion and legislators' vote choice between bills and depending on legislators' vote choice. We address these challenges by estimating a within-between Random Effects (REWB) model (Bell et al., 2019; Bell and Jones, 2015). This model allows us to account for speaker heterogeneity through the estimation of *within* and *between* legislator effects. While within effects are solely estimated with variation of individual speaking behavior across debates, between effects are estimated by comparing speaking behavior between legislators. Furthermore, the model allows us to account for differences between debates through the incorporation of debate-level random intercepts. The following equation formalizes the model:

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

y_{it} is legislator i's speech emphasis in debate t. x_{it} is the alignment between legislator i's vote after debate t and the preference of her electoral district. $\bar{x}_i$ is the average alignment between legislator i's votes and the preference of her district. υ_i are random intercepts for each legislator i and υ_t are random intercepts for each debate t. z_{ki} represent K additional individual-level control variables.

The model has several properties worth noting. Importantly, the 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 vote-alignment from individual legislators' average vote-alignment. The respective coefficient β_{1W} represents the average *within* effect of vote-alignment, that is the expected change in a legislator's speech emphasis caused by changes of preference in his or her electoral district. This makes β_{1W} the main coefficient of interest. Second, the model incorporates legislators' average vote-alignment $(\bar{x}_i)$ as a covariate. The respective 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 support for their vote choices on average. In sum, the within effect captures the consequences of variation in district opinion across debates.

Thus, it indicates whether legislators are responsive to variation in public opinion in their electoral district. In contrast, the between effect captures the consequences of variation of public opinion between electoral districts. Thus, this is less an estimate of individual responsiveness but assesses whether high levels of ideological coherence lead legislators to be more willing to deliver emphatic speeches.

While the separation of the effect of vote-alignment into within and between legislator effects accounts for systematic differences in the speech emphasis of individual legislators, it does not account for the clustering of vote-alignment at the debate level. We address this concern by including random intercepts at the debate level (υ_t). Because there is not only clustering at the debate level but also depending on legislators' vote choice (see figure 3), we estimate separate models for legislators who rise in opposition and legislators who rise in support of a bill.

Our set of control variables comprises five potential confounders. We account for legislators' party affiliation with a dummy for Republican legislators. As legislators who vote against the party line may face pressures not to signal this behavior, we include a dummy 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 score (Lewis et al., 2020). Finally, we control for gender to account for the possibility that male and female legislators differ in their presentational styles.

District preferences and signaling in legislative speech

We estimate four models to test our theoretical expectations. The first two model specifications are estimated for legislators who delivered speeches in opposition of the underlying bill. Model (1) is based on public opinion estimates from MrP while Model (2) uses public opinion estimates from BARP. The estimated effects show whether legislators become more emphatic in their speeches as their electoral district becomes increasingly opposed to the underlying piece of legislation. Model (3) and (4) are based on observations of legislators who rose in support of the underlying bill, with public opinion estimates from MrP and BARP respectively. These models test whether supporting legislators increase the emphasis in their speeches if their electoral district becomes more supportive of the underlying piece of legislation.

The results in Table 5 provide support for the notion that legislators react to public opinion when rising in opposition to a bill. Specifically, legislators become more emphatic as their districts become increasingly hostile to the underlying piece of legislation. Model

Table 3: Estimation Results (1)

	In opposition		In support	
	(1) MrP	(2) BARP	(3) MrP	(4) BARP
Intercept	−0.57	−0.51	−0.86	−0.71
	(0.49)	(0.52)	(0.48)	(0.51)
Vote-District Alignment, *within*	0.76	0.71	−0.24	−0.27
	(0.32)	(0.33)	(0.27)	(0.28)
Vote-District Alignment, *between*	1.86	1.73	0.61	0.33
	(0.68)	(0.80)	(0.53)	(0.58)
Controls	✓	✓	✓	✓
N(Legislators)	294	296	403	409
N(Debates)	25	25	25	25
Var: Legislators (Intercept)	0.21	0.21	0.17	0.18
Var: Debates (Intercept)	0.04	0.04	0.02	0.02
Var: Residual	0.25	0.25	0.24	0.24
AIC	1447.62	1455.54	1617.17	1649.48
BIC	1498.41	1506.36	1669.38	1701.84
Log Likelihood	−712.81	−716.77	−797.59	−813.74
N	748	750	851	863

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

(1) and (2) show that there is a positive *within* effect of vote-district alignment on speech emphasis (0.76 and 0.71 depending on the public opinion estimates). This implies that opposing legislators react to varying public support for bills by putting more emphasis on speeches when the bill is especially unpopular in their district. Figure 4 helps to assess the substantive 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 her district are—like herself—opposed to the debated legislation. In the second debate, 75% are opposed, making her vote more aligned with public opinion in her district.[11] Using the estimate from model (1), in the second debate we would expect this legislator to deliver a speech that scores 0.19 points higher on the emphasis scale compared to a speech during the first debate. This is equivalent to 0.45 standard deviations of the de-meaned distribution of the speech emphasis.

Providing additional support for the argument, both models show a positive *between* effect of vote-district alignment on speech emphasis (1.86 and 1.73 depending on the public opinion estimate). This is additional evidence for an effect of public opinion and the delivery style of legislative speech: Legislators whose opposing vote constantly shows high alignment with their electorate tend to deliver more emphatic speeches compared to legislators with less electoral support for their votes. To assess the substantive meaning

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

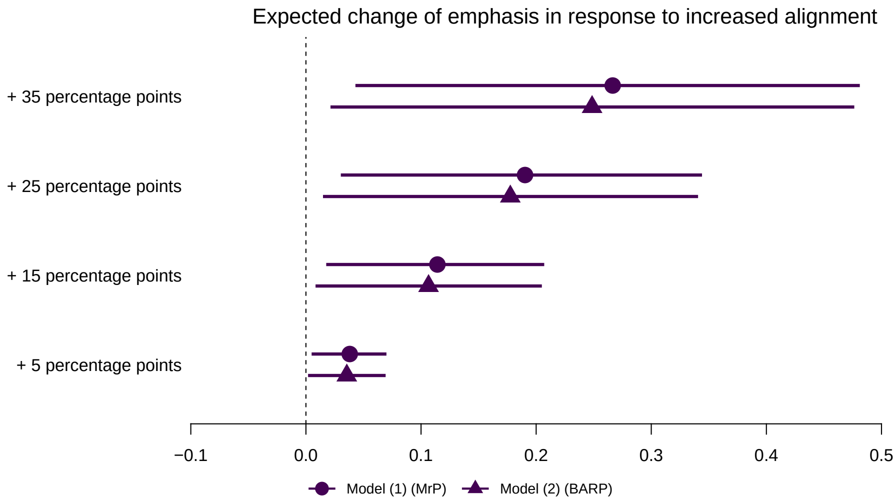

Figure 4: First Differences and 95% Confidence Intervals illustrating the expected change of speech emphasis in response to increased alignment between a legislator's No vote and public opinion in the district.

Note: Simulations are based on model (1) and model (2) in table 3. The baseline value of the de-meaned district alignment is set to −0.28 (minimum when using the MrP estimates), the mean level of vote-alignment is set to the empirical mean (0.47 when using MrP, 0.46 when using BARP), Republican is set to zero, vote with party is set to 1, seniority is set to its mean (15.6), gender is set to zero.

of the coefficients, consider two legislators who are very similar but represent districts whose voters differ in their preferences on key pieces of legislation. This means that the legislators are in the same party, cast the same votes, share the same ideological stances, have similar experiences, and are of the same gender. Yet, legislator A enjoys higher alignment between her votes and public opinion in her district, meaning that more voters in her district take the same stance as her compared to legislator B and voters in her district. Suppose that the difference of vote alignment between legislator A and B amounts to 17 percentage points on average.[12] The models predict that this difference has implications for how legislators A and B present themselves on the floor: Legislator A would deliver speeches that score 0.32 points higher on the emphasis scale on average compared to legislator B. This amounts to 0.55 standard deviations of the mean emphasis scores.

The results of model (3) and (4) show that this finding does not generalize to legislators who rise in support of a bill. While the estimates from the between effects are in the expected direction, neither within nor between estimates from models have substantially meaningful magnitudes and are indistinguishable from zero at reasonable levels of statistical significance. This challenges the proposition that emphatic legislative speech

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

is equally useful to legislators who rise in support and in opposition of a bill.

Conclusion

Automated analyses of audio and video data have recently begun to make their way into political science research (Dietrich, forthcoming; Dietrich et al., 2019a, b; Knox and Lucas, forthcoming). 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 a systematic study of questions beyond the substance of speeches and a greater appreciation of the nonverbal aspects of legislative 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 to 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 legislative 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 delivery of speeches is responsive to public opinion, we rely 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 analysis shows consistent effects of constituency opinion on the delivery of legislative speeches. Across different model specifications, legislators rising in opposition to a partisan bill were found to deliver more emphatic speeches, the more their districts are opposed to a measure.

Despite consistent effects across model specifications, one limitation should be explicitly addressed. We argued that the Cooperative Congressional Election Study is useful for estimating district preferences on congressional roll call votes and that attitudes towards individual 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 these debates, raising the question whether our findings generalize beyond debates on key bills. On the one hand, there is little reason to expect strategic legislators to be willing to signal their positions when they disagree with their constituents. On the other hand, speeches on inconsequential bills might 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 might 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 for district ideology. Such research is greatly simplified by the promises of computer vision where the trained neural network can easily be deployed to study speeches on other bills.

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 towards 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.

Overall, the study of nonverbal characteristics with emerging computer vision tools holds enormous promise for research on legislative behavior. The present contribution constitutes one of the first attempts to systematically trace and explain the nonverbal characteristics of legislative speech with important implications for legislative research. In line with previous research, our findings underscore that legislators are conscious and strategic in their use of legislative speech and that such strategy does not exhaust itself in the substantive aspects of legislative speech.

References

Amsalem, E., Sheafer, T., Walgrave, S., and Loewen, P. J. (2017). Media motivation and elite rhetoric in comparative perspective. *Political Communication*, 34(3):385–403.

Amsalem, E., Zoizner, A., Sheafer, T., Walgrave, S., and Loewen, P. J. (forthcoming). The effect of politicians' personality on their media visibility. *Communication Research*.

Ansolabehere, S. and Jones, P. E. (2010). Constituents' responses to congressional roll-call voting. *American Journal of Political Science*, 54(3):583–597.

Aytar, Y., Vondrick, C., and Torralba, A. (2016). Soundnet: Learning sound representations from unlabeled video. In *Advances in neural information processing systems*, pages 892–900.

Banning, S. and Coleman, R. (2009). Louder than words: A content analysis of presidential candidates' televised nonverbal communication. *Visual Communication Quarterly*, 16(1):4–17.

Baumann, M., Debus, M., and Müller, J. (2015). Convictions and signals in parliamentary speeches: Dáil Éirann debates on abortion in 2001 and 2013. *Irish Political Studies*, 30(2):199–219.

Bell, A., Fairbrother, M., and Jones, K. (2019). Fixed and random effects models: making an informed choice. *Quality & Quantity*, 53(2):1051–1074.

Bell, A. and Jones, K. (2015). Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods*, 3(1):133–153.

Bisbee, J. (2019). Barp: Improving mister p using bayesian additive regression trees. *American Political Science Review*, 113(4):1060–1065.

Bucy, E. 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, J. K., Birk, T., and Pfau, M. (1990). Nonverbal behaviors, persuasion, and credibility. *Human Communication Research*, 17(1):140–169.

Bäck, H. and Debus, M. (2018). Representing the region on the floor: Socioeconomic characteristics of electoral districts and legislative speechmaking. *Parliamentary Affairs*, 71(1):73–102.

Carreira, J. and Zisserman, A. (2017). Quo vadis, action recognition? a new model and the kinetics dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6299–6308.

Chollet, F. and Allaire, J. J. (2018). *Deep Learning with R*. Manning.

Cook, T. E. (1986). House members as newsmakers: The effects of televising congress. *Legislative Studies Quarterly*, 11(2):203–226.

Diermeier, D., Godbout, J.-F., Yu, B., and Kaufmann, S. (2011). Language and ideology in congress. *British Journal of Political Science*, 42(1):31–55.

Dietrich, B. J. (2015). If a picture is worth a thousand words, what is a video worth? In Browning, R. X., editor, *Exploring the C-SPAN archives: Advancing the research agenda*, pages 241–263. Purdue University Press, West Lafayette.

Dietrich, B. J. (forthcoming). Using motion detection to measure social polarization in the u.s. house of representatives. *Political Analysis*.

Dietrich, B. J., Enos, R. D., and Sen, M. (2019a). Emotional arousal predicts voting on the u.s. supreme court. *Political Analysis*, 27(2):237–243.

Dietrich, B. J., Hayes, M., and O'Brien, D. Z. (2019b). Pitch perfect: Vocal pitch and the emotional intensity of congressional speech. *American Political Science Review*, 113(4):941–962.

Dietrich, B. J. and Juelich, C. L. (2018). When presidential candidates voice party issues, does twitter listen? *Journal of Elections, Public Opinion and Parties*, 28(2):208–224.

Downs, A. (1957). *An economic theory of democracy*. Cambridge.

Esser, F. (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):401–28.

Esser, F. and Strömbäck, J., editors (2014). *Mediatization of politics: Understanding the transformation of Western democracies*. Palgrave Macmillan.

Gelman, A. and Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Gelman, A. and Little, T. C. (1997). Poststratification into many categories using hierarchical logistic regression. *Survey Methodology*, 23(2):127–135.

Gershon, S. A. (2012). Press secretaries, journalists, and editors: Shaping local congressional news coverage. *Political Communication*, 29(2):160–183.

Grimmer, J. (2013a). Appropriators not position takers: The distorting effects of electoral incentives on congressional representation. *American Journal of Political Science*, 57(3):624–642.

Grimmer, J. (2013b). *Representational styles in Congress: What legislators say and why it matters*. Cambridge.

Grimmer, J., Westwood, S. J., and Messing, S. (2015). *The impression of influence: Legislator communication, representation, and democratic accountability*. Princeton.

Grose, C. R. and Middlemass, K. M. (2010). Listen to what i say, not how i vote: Congressional support for the president in washington and at home. *Social Science Quarterly*, 91(1):143–167.

Hall, R. L. (1996). *Participation in Congress*. Yale University Press.

Heiss, R., Schmuck, D., and Matthes, J. (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, B. and Rocca, M. S. (2005). Beyond the roll-call arena: The determinants of position taking congress. *Political Research Quarterly*, 58(2):303–316.

Hill, K. Q. and Hurley, P. A. (2002). Symbolic speeches in the us senate and their representational implications. *Journal of Politics*, 64(1):219–231.

Hill, S. J. and Huber, G. A. (2019). On the meaning of survey reports of roll-call "votes". *American Journal of Political Science*, 63(3):611–625.

Huang, X., Acero, A., and Hon, H.-W. (2001). *Spoken language processing: A guide to theory, algorithm, and system development*. Prentice Hall.

Jones, D. E., Doty, S., Grammich, C., Horsch, J. E., Houseal, R., Lynn, m., Marcum, J. P., Sanchagrin, K. M., and Taylor, R. H. (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.

Joo, J., Bucy, E. P., and Seidel, C. (2019). Automated coding of televised leader displays: Detecting nonverbal political behavior with computer vision and deep learning. *International Journal of Communication*, 13:4044–4066.

Kalaf-Hughes, N. (2020). Representation by the minute: The influence of ethnicity, partisanship and district fit on legislator one-minute floor speeches. *Congress and the Presidency*, 47(1):92–129.

Kay, W., Carreira, J., Simonyan, K., Zhang, B., Hillier, C., Vijayanarasimhan, S., Viola, F., Green, T., Back, T., Natsev, P., et al. (2017). The kinetics human action video dataset. *arXiv preprint arXiv:1705.06950*.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Knox, D. and Lucas, C. (forthcoming). A dynamic model of speech for the social sciences. *American Political Science Review*.

Koppensteiner, M. and Grammer, K. (2010). Motion patterns in political speech and their influence on personality ratings. *Journal of Research in Personality*, 44(3):374–379.

Larsson, A. O. (2020). Winning and losing on social media: Comparing viral political posts across platforms. *Convergence*, 26:639–57.

Lauderdale, B. E. and Herzog, A. (2016). Measuring political positions from legislative speech. *Political Analysis*, 24(3):374–94.

Lax, J. R. and Phillips, J. H. (2009). How should we estimate public opinion in the states? *American Journal of Political Science*, 53(1):107–121.

Lewis, J. B., DeVine, B., Pitcher, L., and Martis, K. C. (2013). Digital boundary definitions of united states congressional districts, 1789-2012.

Lewis, J. B., Poole, K., Rosenthal, H., Boche, A., Rudkin, A., and Sonnet, L. (2020). Voteview: congressional roll-call votes database (2020). *URl: https://voteview. com*.

Lipinski, D. and Neddenriep, G. (2004). Using "new" media to get "old" media coverage: How members of congress utilize their web sites to court journalists. *International Journal of Press/Politics*, 9(1):7–21.

Maier, J., Hampe, J. F., and Jahn, N. (2016). Breaking out of the lab: Measuring real-time response to televised political content in real-world settings. *Public Opinion Quarterly*, 80(2):542–553.

Maier, J., Maurer, M., Reinemann, C., and Faas, T. (2006). Reliability and validity of real-time response measurement: A comparison of two studies of a televised debate in germany. *International Journal of Public Opinion Research*, 19(1):53–73.

Maltzman, F. and Sigelman, L. (1996). The politics of talk: Unconstrained floor time in the us house of representatives. *The Journal of Politics*, 58(3):819–830.

Manson, S., Schroeder, J., Riper, D. V., Kugler, T., and Ruggles, S. (2020). Ipums national historical geographic information system: Version 15.0.

Masters, R. D. and Sullivan, D. G. (1989). Nonverbal displays and political leadership in france and the united states. *Political Behavior*, 11(2):123–156.

Mayhew, D. R. (1974). *Congress: The electoral connection*. Yale University Press.

Monroe, B. L., Colaresi, M. P., and Quinn, K. M. (2008). Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*, 16(4):372–403.

Morris, J. S. (2001). Reexamining the politics of talk: Partisan rhetoric in the 104th house. *Legislative Studies Quarterly*, 26(1):101–121.

Nave, N. N., Shirman, L., and Tenenboim-Weinblatt, K. (2018). Talking it personally: Features of successful political posts on facebook. *Social Media and Society*.

Negrine, R. and Lilleker, D. G. (2002). The professionalization of political communication: Continuities and change in media practices. *European Journal of Communication*, 17(3):305–23.

Nokken, T. P. (2000). Dynamics of congressional loyalty: Party defection and roll-call behavior, 1947-97. *Legislative Studies Quarterly*, 25(3):417–444.

Proksch, S.-O. and Slapin, J. B. (2009). Position taking in european parliament speeches. *British Journal of Political Science*, 40(3):587–611.

Proksch, S.-O. and Slapin, J. B. (2012). Institutional foundations of legislative speech. *American Journal of Political Science*, 56(3):520–537.

Proksch, S.-O. and Slapin, J. B. (2014). *The politics of parliamentary debate*. Cambridge University Press.

Qiu, Z., Yao, T., and Mei, T. (2017). Learning spatio-temporal representation with pseudo-3d residual networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5533–5541.

Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., and Radev, D. R. (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, 54(1):209–228.

Sellers, P. J. and Schaffner, B. F. (2007). Winnnig coverage in the u.s. senate. *Political Communication*, 24(4):377–391.

Sheafer, T. (2001). Charismatic skill and media legitimacy: An actor-centered approach to understanding the political communication competition. *Communication Research*, 28(6):711–736.

Sheafer, T. (2008). Charismatic communication skill, media legitimacy, and electoral success. *Journal of Political Marketing*, 7(1):1–24.

Sheafer, T. and Wolfsfeld, G. (2004). Production assets, news opportunities and publicity for legislators: A study of israeli knesset members. *Legislative Studies Quarterly*, 29(4):611–630.

Squire, P. (1988). Who gets national news coverage in the u.s. senate? *American Politics Quarterly*, 16(2):139–156.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958.

Strömbäck, J. (2008). Four phases of mediatization: An analysis of the mediatization of politics. *International Journal of Press/Politics*, 13(3):228–46.

Taylor, L. and Nitschke, G. (2018). Improving deep learning with generic data augmentation. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1542–1547.

Umit, R. and Auel, K. (2020). Divergent preferences and legislative speeches on brexit. *Journal of Elections, Public Opinion and Parties*, 30(2):202–20.

Vos, D. (2014). Which politicians pass the news gates and why? explaining inconsistencies in research on news coverage of individual politicians. *International Journal of Communication*, 8:2438–2461.

Warshaw, C. and Rodden, J. (2012). How should we measure district-level public opinion on individual issues? *The Journal of Politics*, 74(1):203–219.

Wasike, B. (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, G. and Sheafer, T. (2006). Competing actors and the construction of political news: The contest over waves in israel. *Political Communication*, 23(3):333–354.

Appendices

A Rules for cutting videos

This section documents the rules for manually cutting the videos into sequences that contain material for measuring emphasis in legislative speech. A speech begins after the Speaker of the House 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 speaking member for the first time. This can happen after the legislator begins delivering his or her 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 emphasis in the speech. 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 one of the floor managers grants additional time to a legislator or because the camera shows someone other than the speaker, the sequence ends and a new sequence beings 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 has multiple sequences, all sequences belonging to the same speech are merged.

B Codebook for 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, weak use of hands	Monotone, low voice
Very low	-3	Gaze down, reading, closed posture, no body movement	Notable pauses in speech, low voice

Table 4: Manual speech emphasis coding scheme

C Matching legislative debates, bills, and CCES items

Table 5: CCES items matched with bills, legislative debates and House roll calls

Wave	Item	Bill	Congress	Title	House Vote (Yae-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	04/05/2017
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

D Measuring district level preferences

The key independent variable is the distance between legislators' roll call votes and their districts' median voter preference on the respective bill. We employ two approaches to estimate district level preferences on the bills listed in Table 5: multilevel regression and poststratification (MrP) (Gelman and Little, 1997; Warshaw and Rodden, 2012) and Bayesian additive regression trees and 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 with information on respondents' preferences, their district, and demographics. For this, we draw on data from multiple waves of the Cooperative Congressional Election Study (CCES). The CCES common content includes questions on preferences towards specific bills (dependent variables of the multilevel model) as well as information about respondents' congressional district and demographics (independent variables). (2) To estimate district preferences using poststratification, we use census data with information on the joint distribution of demographic and geographic information of voters in each congressional district.[13] This data comes from the US Census Bureau. Specifically, we draw on the American Community Survey (ACS). All district-level data sets for poststratification are listed in Table 6.

D.1 Multilevel regression and poststratification

The estimation procedure closely follows Warshaw and Rodden (2012). We measure individual preferences towards specific bills using the "roll-call" items in the CCES. Table 5 reports how we match individual bills to specific items in the CCES. 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. On the district level, we include respondents' state, median household income in the district, percentage of veterans, the natural log of the population density[14], and the share of same-sex marriages. On the state level, we further include presidential vote shares in the past presidential election and the percentage of evangelical Protestants as well as Mormons. The latter data comes from the Religious Congregations and Membership Study (Jones et al., 2002). We incorporate this information in the hierarchical model as follows:

$$Pr(y_i = 1) = logit^{-1}(\gamma_0 + \alpha^{\text{race}}_{r[i]} + \alpha^{\text{gender}}_{g[i]} + \alpha^{\text{educ}}_{e[i]} + \alpha^{\text{district}}_{d[i]}) \tag{3}$$

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

[14] We use shapefiles and population estimates to estimate the population density. Shapefiles come from Lewis et al. (2013) and the United States Census Bureau.

where

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

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

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

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

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

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

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

First, we estimate the model for each key vote. Second, we build a 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 all possible 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 all possible combinations of characteristics weighted by the true share of residents in the district with the respective combination of characteristics. This yields an estimate of the district preference towards all bills in the sample ranging from zero to one, where low values indicate opposition to the bill and high values indicate support.

D.2 Bayesian additive regression trees and 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.[15] For estimation, we rely on the R-Package `BARP` (Bisbee, 2019).

Estimation results

Figure 5 depicts the distributions of the resulting district preference estimates by bill. Figure 6 shows the bivariate distribution of the MrP and BARP estimates. To assess the validity of the estimated district preferences, we correlate the estimates with the district-level Republican vote share in the previous Congressional election. The results displayed in Figure 7 suggest a good performance of the MrP models. We observe high correlations between the estimates and the district-level vote shares for all partisan pieces of legislation.

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

Table 6: 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

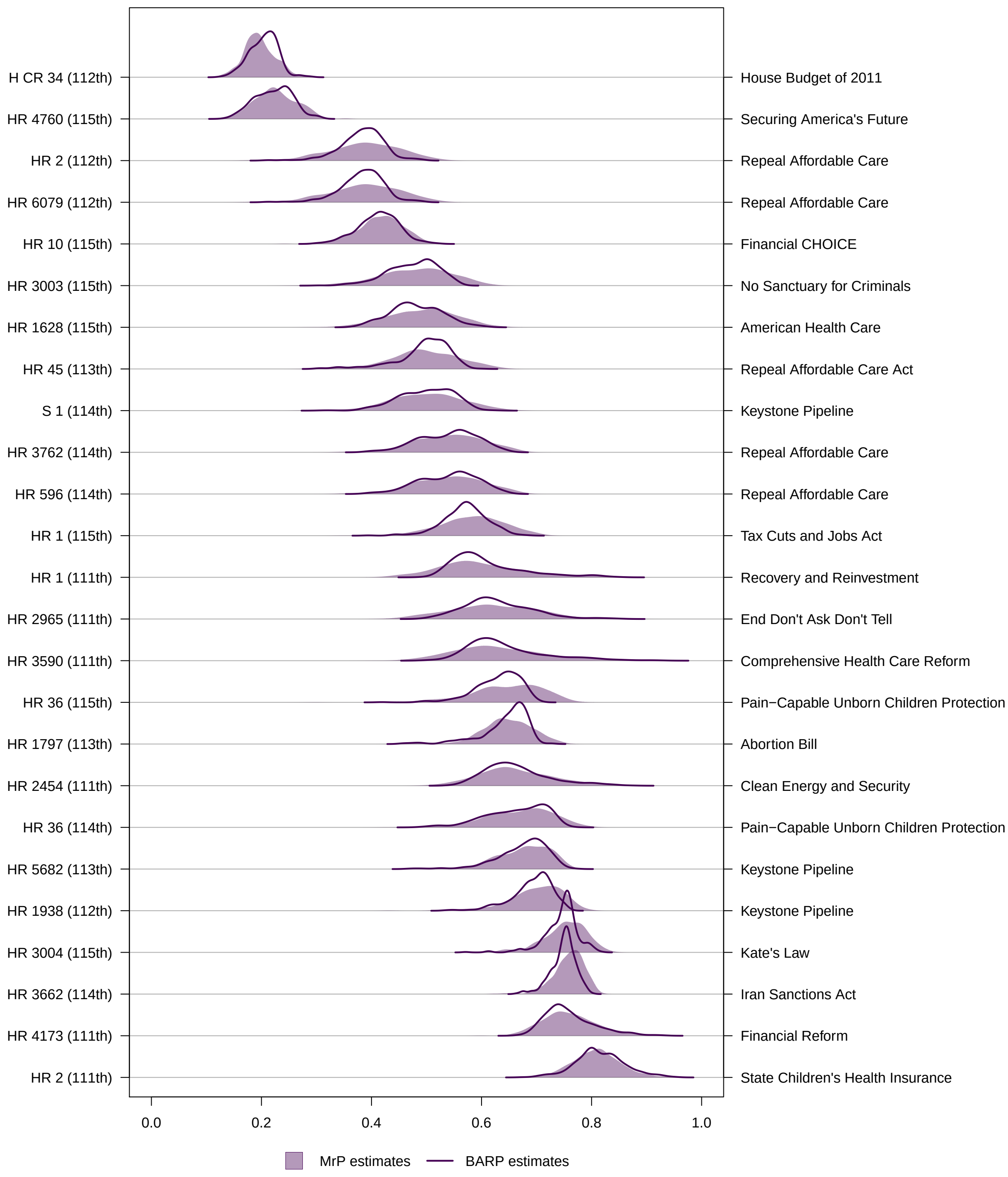

Figure 5: Distributions of district median voter preference estimates

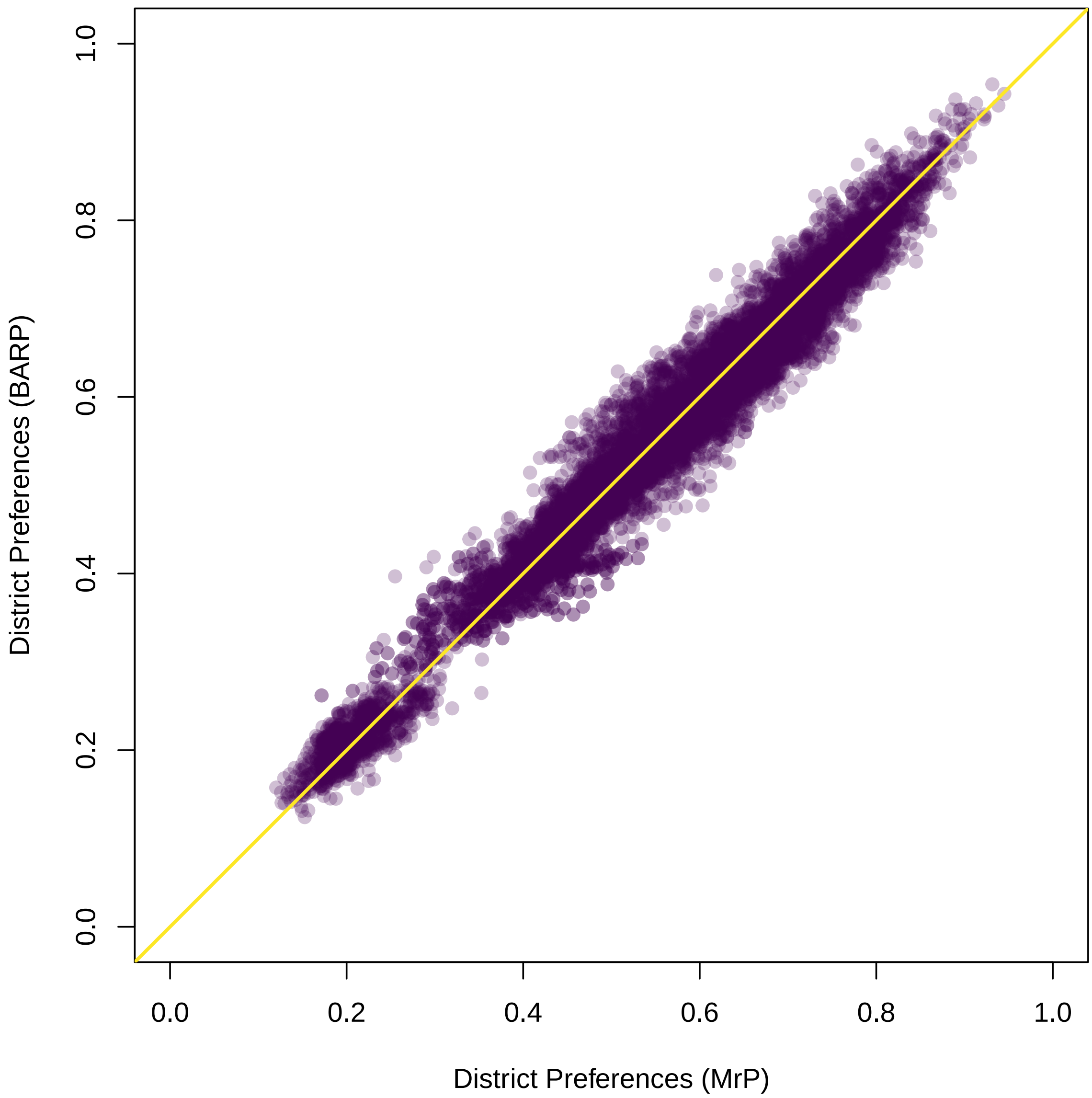

Figure 6: Scatterplot of MrP and BARP district preference estimates.

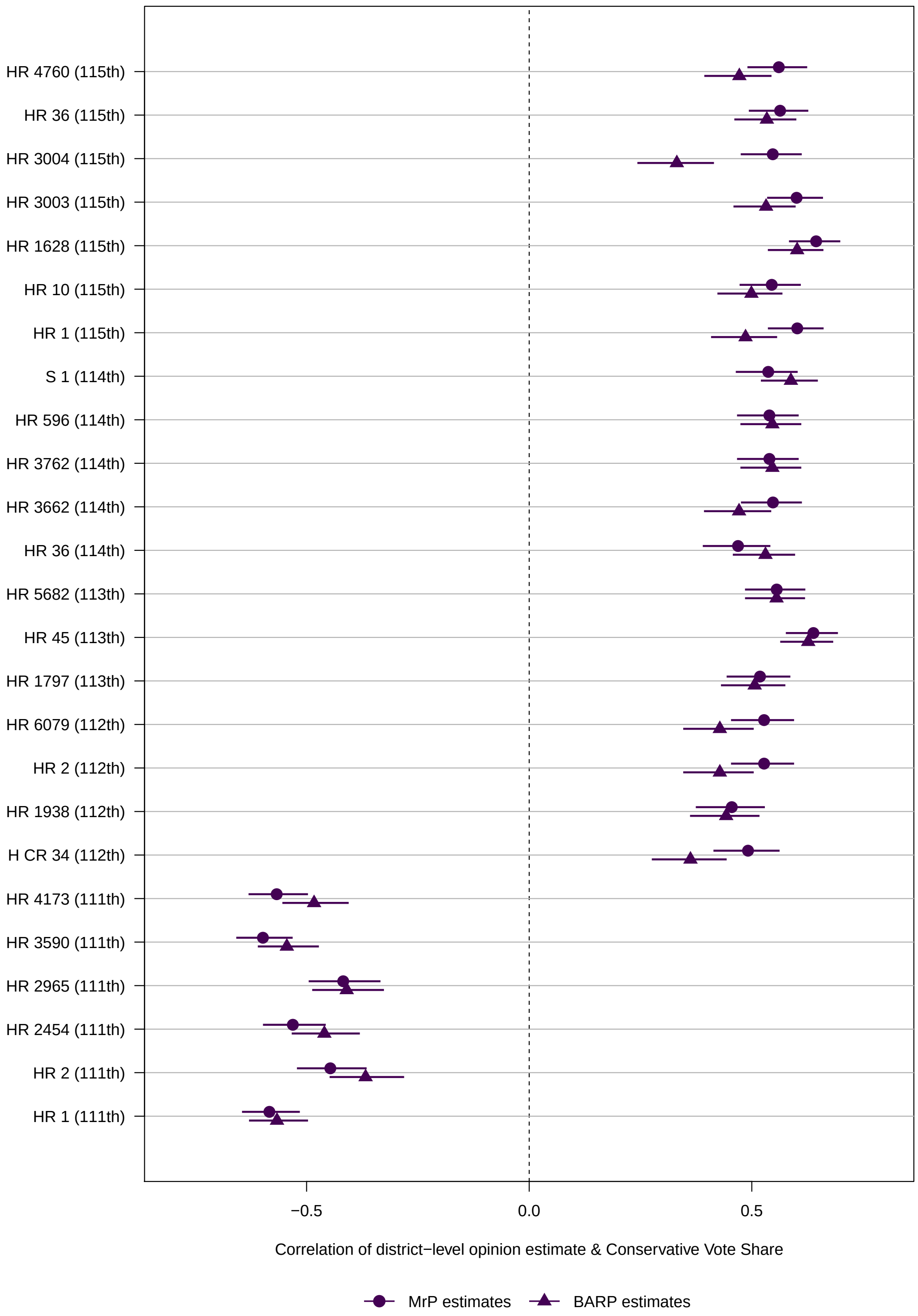

Figure 7: Correlation of MrP and BARP estimates with district-level conservative vote share by legislation.

Note: Error bars indicate 95%-confidence intervals.

E Robustness

E.1 Sensitivity to computation of the dependent variable

Figure 8 shows within and between effect coefficient estimates for the effect of vote-alignment on speech emphasis based on different computations of the speech emphasis variable.

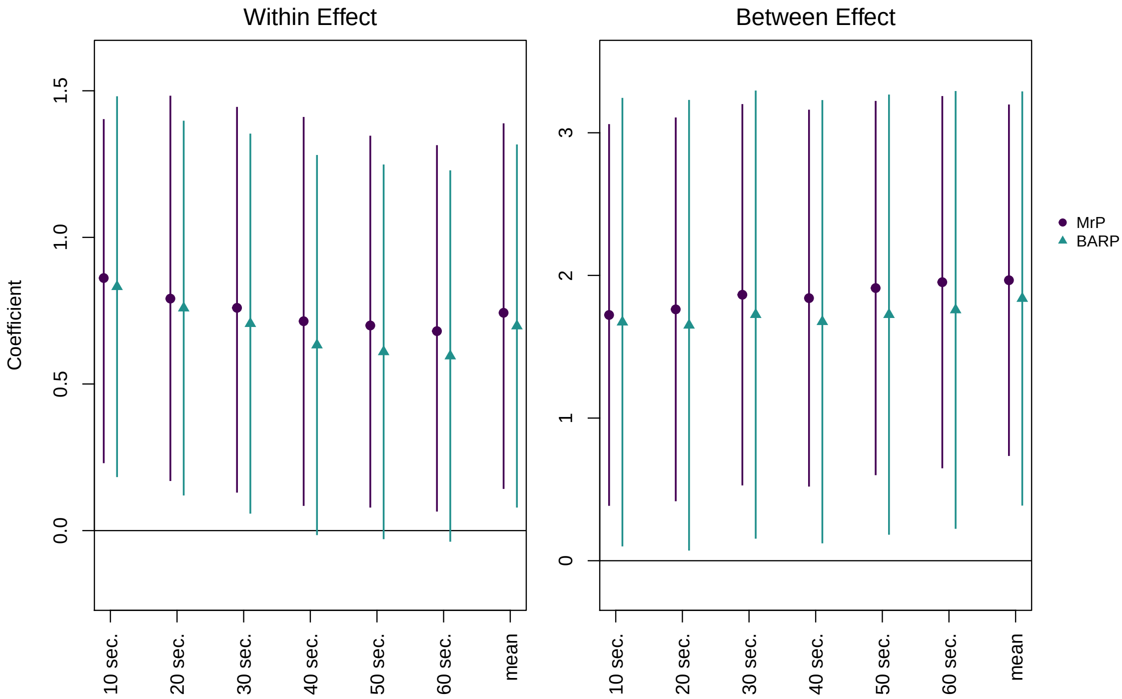

Figure 8: Point estimates and 95% Confidence Intervals of between effects of vote-alignment based on different computations of the dependent variable (speech emphasis)