Appendix

Table A1: Full Congressional OLS Model, TBIP

Independent Variable	*Dependent variable:*		
	Vote	Speech	Tweet
Constant	-1.171*** (0.128)	-0.657 (0.613)	-2.030*** (0.353)
House Candidate Vote Share, 2016	-0.003*** (0.001)	0.004 (0.004)	0.006* (0.003)
House Candidate Vote Share, 2018	0.001 (0.001)	-0.002 (0.005)	-0.0002 (0.003)
Party, Republican	1.829*** (0.179)	2.193** (0.857)	2.376*** (0.495)
Trump Vote Share, 2016	0.005*** (0.001)	0.008 (0.007)	0.024*** (0.004)
District percent white	0.0002 (0.001)	-0.001 (0.003)	-0.001 (0.001)
District unemployment rate	0.011 (0.011)	0.009 (0.054)	0.018 (0.031)
Urban	0.048 (0.032)	-0.120 (0.152)	-0.083 (0.088)
Suburban	0.007 (0.026)	-0.217* (0.124)	-0.066 (0.071)
Number of House Terms	-0.003 (0.001)	-0.041*** (0.008)	0.023*** (0.004)
MC Gender	-0.011 (0.018)	-0.075 (0.089)	-0.115** (0.052))
Problem Solvers Caucus	0.007 (0.037)	0.071 (0.071)	-0.048 (0.096)
Progressive Caucus	-0.026 (0.020)	-0.161* (0.095)	-0.145*** (0.055)
New Democratic Coalition	0.035* (0.021)	0.050 (0.099)	-0.004 (0.084)
Blue Dog Coalition	0.294*** (0.031)	0.561*** (0.147)	0.568*** (0.057)
RSC	0.139*** (0.022)	0.188* (0.106)	0.148** (0.062)
Freedom Caucus	0.069*** (0.023)	-0.141 (0.112)	-0.025 (0.065)
Democratic Party leader	-0.042 (0.028)	-0.218 (0.133)	-0.148* (0.076)
GOP Party leader	0.005 (0.036)	-0.325* (0.172)	-0.027 (0.099)
Member of top committee	-0.006 (0.028)	-0.404*** (0.103)	-0.081 (0.059)
Committee Chair	0.038 (0.035)	0.281* (0.168)	-0.232** (0.097)

House Candidate Vote Share, 2016*Republican	0.004** (0.002)	-0.002 (0.008)	-0.001 (0.004)
House Candidate Vote Share, 2018*Republican	0.001 (0.002)	-0.004 (0.011)	-0.004 (0.006)
Trump Vote Share, 2016*Republican	-0.002 (0.001)	-0.006 (0.011)	-0.015** (0.003)
District percent white*Republican	-0.003** (0.002)	-0.006 (0.005)	-0.003 (0.003)
District unemployment rate*Republican	0.011 (0.019)	0.057 (0.095)	0.028 (0.055)
Urban district*Republican	-0.006 (0.117)	-0.381 (0.560)	-1.446*** (0.321)
Suburban District*Republican	0.006 (0.006)	-0.013 (0.162)	-0.011 (0.094)
MC Gender*Republican	0.002 (0.034)	-0.076 (0.185)	0.138 (0.107)
Problem Solvers*Republican	-0.059 (0.039)	-0.255 (0.184)	0.007 (0.109)
Top Committee*Republican	-0.005 (0.033)	-0.353** (0.155)	0.040 (0.089)
Committee Chair*Republican	-0.036 (0.053)	-0.325 (0.025)	0.325** (0.145)
R^2	0.991	0.768	0.930

Note: Party (Republican) acts as an interaction variable for those relevant variables (any independent variable that applies to both Democratic and Republican members). $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table A2: Changes in Ranked Ideal Point Position, 115th – 116th Congress Interactive Models

Independent Variable	*Dependent Variable*	
	Shift Vote to Speech	Shift Vote to Tweet
Republican MC	92.184 (100.61)	191.165** (92.342)
Female MC	18.709 (11.673)	15.573* (8.848)
Republican * Female MC	-30.956 (19.699)	-5.249 (22.111)
Non-White MC	6.379 (14.544)	17.029* (10.428)
Republican * Non-White MC	-18.972 (26.659)	42.126 (34.843)
Number of terms served	-3.062** (1.529)	-3.739*** (1.365)
Republican * Number of terms served	6.609*** (1.989)	-2.686 (2.017)
Percent vote in last election	0.337 (0.534)	-.433 (0.503)
Republican * Percent vote last election	0.936 (.848)	1.840* (1.054)
District vote for Trump	0.452	-0.396

	(0.970)	(0.761)
Republican * District vote for Trump	-0.140	1.034
	(1.344)	(1.336)
Suburban district	50.364***	25.325*
	(16.577)	(14.534)
Republican * Suburban district	-18.444	-18.454
	(21.248)	(19.591)
Urban district	32.234	40.091**
	(21.329)	(17.173)
Republican * Urban district	17.271	41.765
	(31.969)	(61.644)
District percent white	0.336	0.095
	(0.455)	(0.395)
Republican * District percent white	-1.367*	-2.225***
	(0.716)	(0.703)
District unemployment rate	-12.128	-6.720
	(7.510)	(6.772)
Republican * District unemployment	15.138*	-29.687***
	(9.013)	(8.992)
Problem Solvers Caucus	-28.163**	-25.213***
	(14.002)	(10.185)
Republican * Problem Solvers Caucus	12.493	-7.128
	(21.300)	(18.965)
Republican Study Group (Rep)	24.492**	9.782
	(11.885)	(13.960)
Freedom Caucus (Rep)	73.036***	29.549*
	(15.839)	(16.720)
Progressive Caucus (Dem)	21.389*	14.242
	(12.584)	(11.233)
New Democratic Coalition (Dem)	7.381	13.615
	(14.164)	(11.511)
Blue Dog Coalition (Dem)	4.293	7.683
	(18.828)	(9.534)
Party leader	11.264	10.630
	(18.253)	(17.528)
Republican * Party leader	25.750	-27.521*
	(19.332)	(14.772)
Committee chair	-40.698*	0.741
	(24.261)	(23.652)
Republican * Committee chair	55.491*	-26.406
	(32.748)	(28.826)
Top committee member	10.535	5.448
	(10.580)	(8.627)
Republican * Top committee	-7.748	-50.197***
	(16.206)	(14.648)
Constant	-58.703	30.130
	(73.762)	(62.414)
Observations	489	471
R^2	0.268	0.349

*Note: This includes members from the 115th and 116th Congress. Estimated with OLS regression. *p<0.1; **p<0.05; ***p<0.01. Negative sign indicates more conservative than vote, positive sign indicates more conservative than vote*

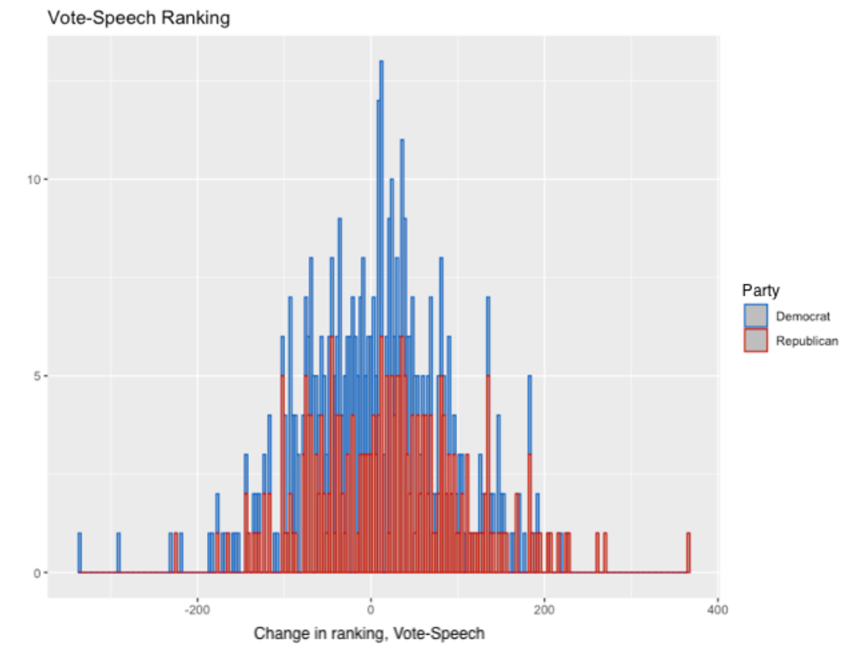

Figure A1: Independent-level shift from vote-based ideal points to floor-speech-based ideal points, 115th and 116th Congress

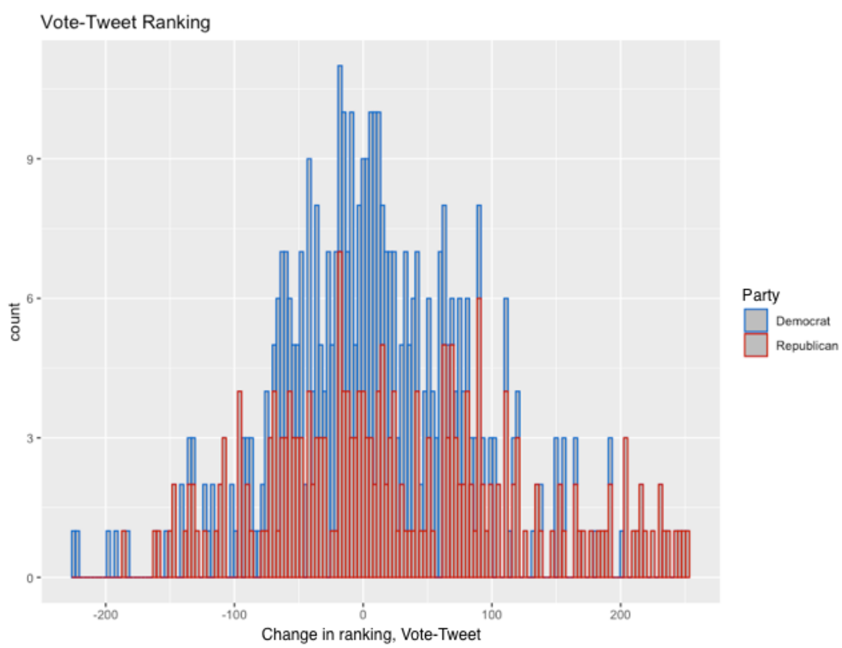

Figure A2: Independent-level shift from vote-based ideal points to Tweet-based ideal points, 115th and 116th Congress

Preprocessing details

For the most part, we followed the text processing settings described in Appendix B in Vafa et al. (2020), and some major differences were already described in the main paper in “Preparing Text Data.” In case of floor speeches, speakers giving less than 25 speeches were removed, and terms used by less than 50 representatives in the data were also removed when processing and finalizing the vocabulary.

We also remove procedural speeches - short speech segments containing pro forma language alone such as “I yield back the balance of my time.” following our own implementation of the procedure to identify such speeches laid out in Card et al. (2022). Their procedure works with the assumption that “short speeches which are repeated 128 verbatim or nearly verbatim many times should be treated as procedural.” — after some manual and heuristic based identification of a few positive and negative samples, a regularized logistic regression classifier was able to identify such procedural speeches with about 99% accuracy. We refer the reader to Card et al., 2022 for details (“identifying procedural speeches” in the supplementary information). Around ~22k speech texts out of an initial ~81k speeches were removed as procedural speeches due to this step.

The speeches containing only pro forma language can be misleading and negatively affect ideal point modeling since they would not demonstrate partisan language. We subjectively find that removing these speeches improves the partisan topics discovered and recommend this step for application of TBIP in future work.

In case of tweets, terms used by less than 5 representatives in the data were also removed when processing and finalizing the vocabulary. In both datasets, unigrams, bigrams, and trigrams were allowed.

Our final processed dataset for floor speeches contains 58,932 speeches and 11,433 unique terms in the vocabulary. For Twitter data, our final processed dataset contains 295,327 tweets and 9,343 unique terms in the vocabulary.
For both datasets, we run the TBIP model with 50 topics.

Assessing model reliability

Given a set of initial topics (estimated in the first step) and for a given dataset, we assess the stability of the ideal point values for the authors (legislators) estimated by TBIP by running the model 10 times, and then for all pairs of runs, we look at the correlation between the ideal point estimates of the authors. This is done because some variables in the estimation process are randomly initialized and it is important to know how much of a variance it can cause in the final TBIP estimates the model outputs. This allows us to judge if one can reliably go ahead and use the model results of a single run.

The ideal point estimates learned on multiple runs (with different random seed) are highly correlated across pairs of runs (average Spearman $r = 0.675, p < 0.0005$ and average Pearson $r =$

0.713, p < 0.0005 for floor speeches; average Spearman $r = 0.897$, $p < 0.0005$ and average Pearson $r = 0.908$, $p < 0.0005$ for tweets).

Note that we did the above stability analysis with the original Poisson estimation-based topic initialization, but we expect the stability results to be unaffected by initializing with LDA as described above since ideal point estimates correlate highly (average Pearson $r > 0.9$) when comparing Poisson vs MALLET initialization (though topics discovered are subjectively determined to be more coherent). The TBIP model appears to be quite stable in terms of ideal point values it discovers for authors.

Annotator Process and Results

Four subject-matter experts (PhD graduate students) separately analyzed the coherence and polarization of the TBIP topics. Below is an overview of the directions given to the annotators and the results of their analysis.

Coherence: Rate the topic, on a 1-3 scale, on its coherence (1 being not coherent, 3 being very coherent). Does that topic represent an easily identifiable category or a meaningful concept? The top words for the topic, along with the top documents associated with that topic help make this judgment. Broadly, a set of items can be said to be *coherent* if they enable human recognition of an identifiable category when viewed together.

Coherence Rating	Floor Speeches		Tweets	
	Annotator 1	Annotator 2	Annotator 1	Annotator 2
1.0	4%	16%	0%	22%
2.0	2%	6%	2%	8%
3.0	94%	78%	98%	70%
Cohen's Kappa inter-annotator agreement	0.151		0.040	

Polarization: If the topic is rated above 1 for coherence, rate the topic, on a 1-3 scale (1 being not polarized, 3 being polarized), on the expected polarization of this topic: do you expect meaningful ideological differences in the way liberals and conservatives would talk about the category or concept or issue or the stance they would hold on that issue? The top words for the topic, along with the top documents associated with that topic, as well as your personal knowledge of American politics will help make this judgment.

Polarization Rating	Floor Speeches		Tweets	
	Annotator 1	Annotator 2	Annotator 1	Annotator 2

1.0	44%	46%	48%	52%
2.0	0%	8%	0%	10%
3.0	52%	30%	52%	16%
N/A	4%	16%	0%	22%
Cohen's Kappa inter-annotator agreement	0.370		0.221	

Based on the topic labels after consensus between the two annotators (in the pre-TBIP stage), we filtered topics that were not coherent or represented non-substantive issues, which left us with 42 topics for speeches and 40 topics for tweets that were then presented to two different annotators in the post-TBIP annotation stage.

Label Applicability to Perspectives: Looking at the top words shown for the perspective a) of a particular issue number (columns A and B under the issue number), select one of the three options based on your judgment of whether the Issue Label is a suitable label for the underlying concept or category you think the words shown for "a)" represent. Put differently, is the shown <Issue Label> applicable to the words shown for the words under "a)" for that issue?

Perspective a)	Floor Speeches		Tweets	
	Annotator 1	Annotator 2	Annotator 1	Annotator 2
IS about <Issue Label>	90.48%	83.33%	92.5%	57.5%
MIGHT be about <Issue Label>	7.14%	14.29%	7.5%	40.0%
IS NOT about <Issue Label>	2.38%	2.38%	0.0%	2.5%
Cohen's Kappa inter-annotator agreement	0.806		0.594	

Perspective b)	Floor Speeches		Tweets	
	Annotator 1	Annotator 2	Annotator 1	Annotator 2
IS about <Issue Label>	90.48%	80.95%	75.0%	65.0%
MIGHT be about <Issue Label>	9.52%	14.29%	22.5%	30.0%
IS NOT about <Issue Label>	0.0%	4.76%	2.5%	5.0%
Cohen's Kappa inter-annotator agreement	0.830		0.619	

Ideological Polarization

a) and b) represent polarized perspectives for the issue?	Floor Speeches		Tweets	
	Annotator 1	Annotator 2	Annotator 1	Annotator 2
YES	31.71%	35.9%	28.21%	42.11%
SOMEWHAT	41.46%	30.77%	28.21%	31.58%
NO	24.39%	17.95%	43.59%	15.79%
UNSURE	2.44%	15.38%	0.0%	10.53%
Cohen's Kappa inter-annotator agreement	0.330		0.435	

For floor speeches, 81.82% of initial topics (before any ideal point modeling) expected to have political polarization by either annotator were deemed at least somewhat ideologically polarized when looking at the polarized perspectives (given by TBIP) by either annotator.

For Tweets, 72.73% of initial topics (before any ideal point modeling) expected to have political polarization by either annotator were deemed at least somewhat ideologically polarized when looking at the polarized perspectives (given by TBIP) by either annotator.

For floor speeches, for the issues that the annotator deemed ideologically polarized when looking at the two perspectives given by TBIP:

- For issue perspective that were the liberal perspective per TBIP, annotator 1 identified 90.6% of them as liberal, and they were unsure about the remaining 9.4%.
- For issue perspective that were the conservative perspective per TBIP, annotator 1 identified 78.1% of them as conservative, and they were unsure about the remaining 21.9%.
- For issue perspective that were the liberal perspective per TBIP, annotator 2 identified 67.7% of them as liberal, 8.8% as conservative, and they were unsure about the remaining 24%.
- For issue perspective that were the conservative perspective per TBIP, annotator 2 identified 67.7% of them as conservative, 8.8% as liberal, and they were unsure about the remaining 24%.

For tweets, for the issues that the annotator deemed ideologically polarized when looking at the two perspectives given by TBIP:

- For issue perspective that were the liberal perspective per TBIP, annotator 1 identified 96.7% of them as liberal, and they were unsure about the remaining 3.3%.
- For issue perspective that were the conservative perspective per TBIP, annotator 1 identified 76.7% of them as conservative, 6.7% as liberal, and they were unsure about the remaining 16.6%.

- For issue perspective that were the liberal perspective per TBIP, annotator 2 identified 84.9% of them as liberal, and they were unsure about the remaining 15.1%.
- For issue perspective that were the conservative perspective per TBIP, annotator 2 identified 84.9% of them as conservative, and they were unsure about the remaining 15.1%.