

Analyzing Candidates' Ideological Messaging Throughout the Electoral Cycle

Benjamin Wittenbrink

Advised by Prof. Matthew Gentzkow



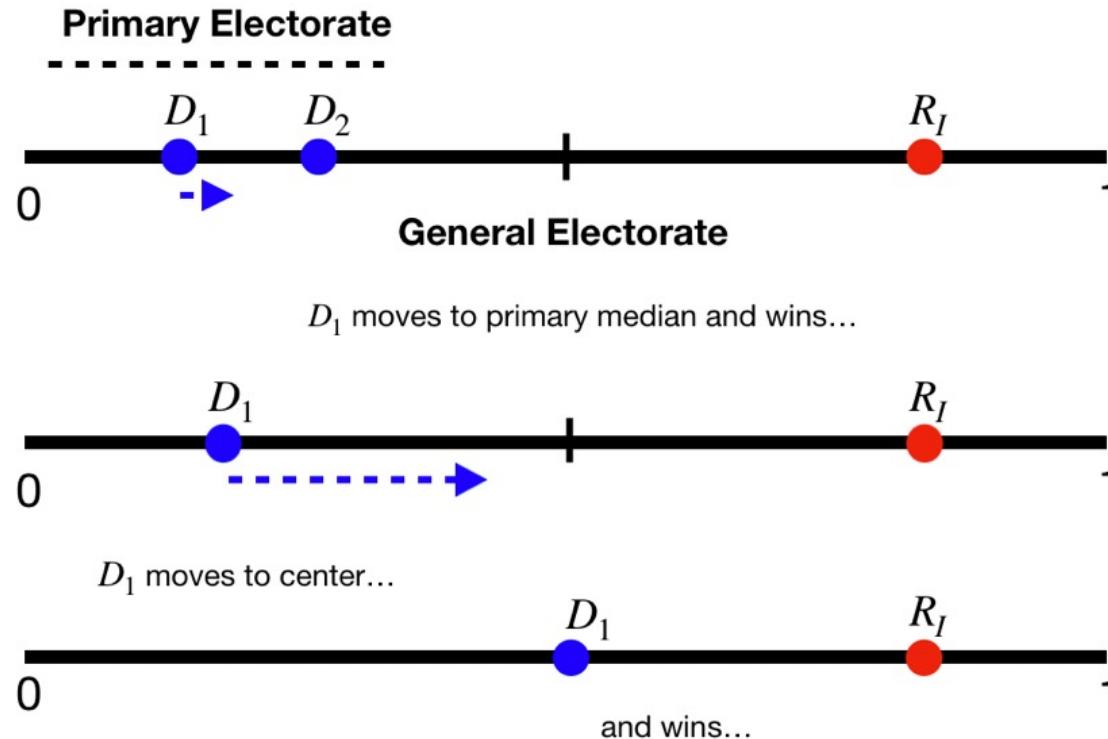
Question:

**How do candidates strategically
manipulate their ideological
rhetoric over the election cycle?**

“Run to the **right in the primaries,
then run to the **center** in the
general election.”**

– Richard Nixon

Intuition:



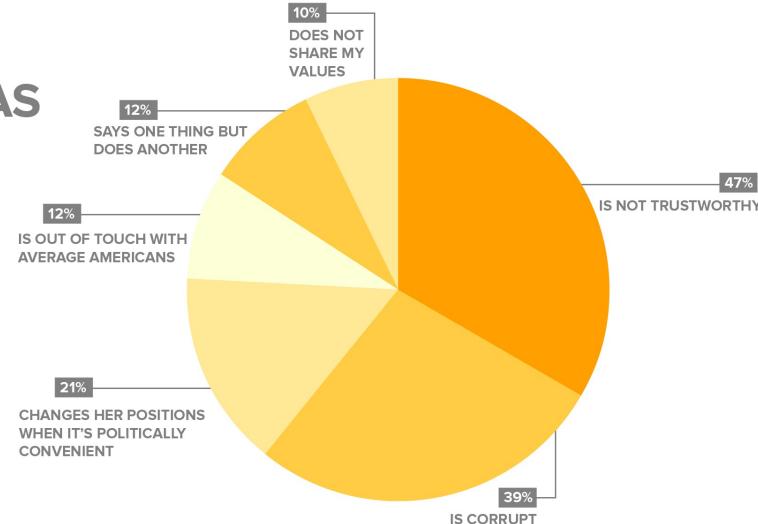
MORNING CONSULT

HILLARY CLINTON HAS A TRUST PROBLEM...

Morning Consult asked voters with an unfavorable view of either Hillary Clinton or Donald Trump to pick the top two reasons they dislike the candidate from a list of more than a dozen negative attributes.

Clinton's unfavorable ratings are clearly driven by views that she is not trustworthy and corrupt, reasons cited by more than a third of voters.

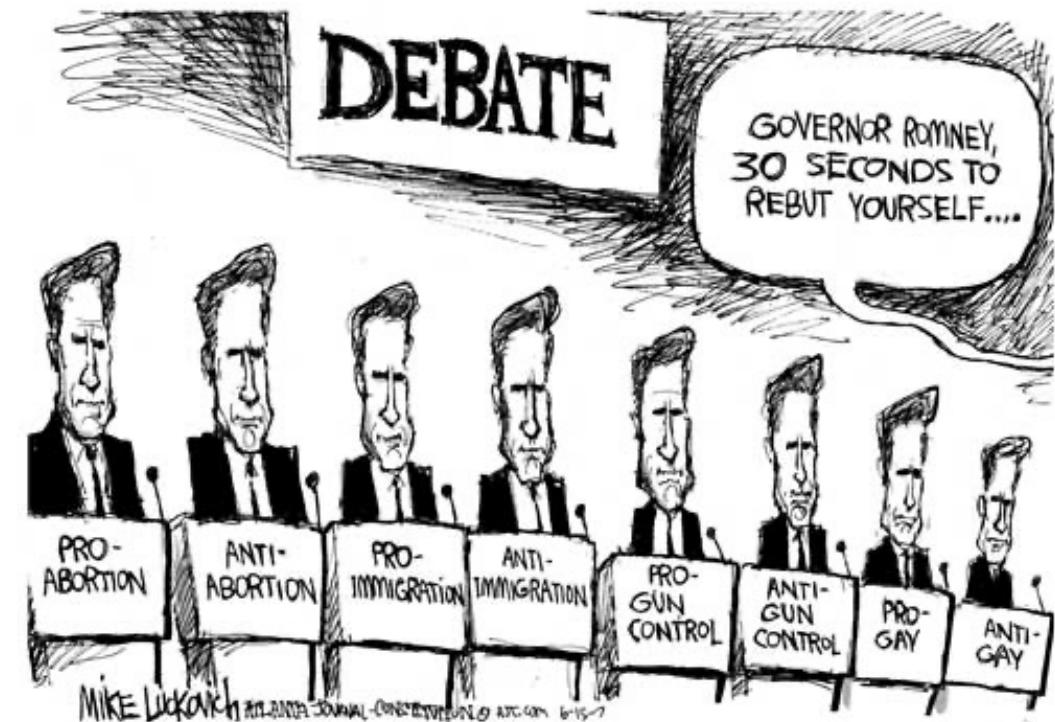
*Numbers do not add up to 100% because survey respondents were asked to pick their top two reasons



FOR MORE, VISIT MORNINGCONSULTINTELLIGENCE.COM



Introducing the Flip-O-Meter



Source from left to right, top to bottom: Morning Consult; History News; PolitiFact; Atlanta-Journal

Hypotheses

1. Congressional candidates should moderate from the primary to the general election.
2. The extent of moderation among congressional candidates in races with competitive general elections should exceed that of those in uncompetitive races.
3. Moderation among incumbent candidates should be less than among non-incumbents.

Contributions

- Political science literature contains little empirical information on the evolution of candidate positions
 - My focus on text data expands the available candidate pool; captures candidates in direct communication with voters; is high-frequency yielding a continuous measure
 - I provide a generalized methodological approach to measuring ideology over time for all candidates
- I provide one of the first empirical analyses on post-primary moderation hypothesis
 - Important theoretical and policy implications, particularly in times of growing polarization among the electorate and Congress

Data

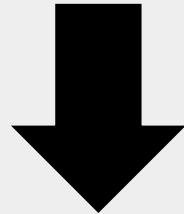


Data

- Two samples:
 - Baseline: members from 116th Congress (House: train, Senate: validation)
 - Candidate: all congressional candidates in 2020 election cycle
- "Gold-Standard Label" for baseline sample: DW-Nominate 1 scores based on roll-call voting
- Various sources to obtain member/candidate- and race-specific metadata
- Twitter API to obtain text data for both samples

Methodology

Extracting ideologically meaningful,
quantitative estimates from
natural language



Alexandria Ocasio-Cortez @AOC

...

The nerve of people who ask “how are you going to pay for it?” whenever we propose building advanced public education, healthcare, & climate infrastructure yet defend a system where Trump pays \$750 in taxes and Amazon pays none is beyond me.

11:06 AM · Sep 29, 2020 · Twitter for iPhone

DW-NOM: -0.47

Methodological Approaches

1. Multinomial Inverse Regression (MNIR):

- Data-driven approach to select most partisan bigrams; use the occurrences of these bigrams to specify a multinomial model of speech

2. Moral Foundations Dictionary (MFD):

- Theoretically-derived dictionary of keywords to construct a measure of the frequency with which candidates invoke moral values associated with political convictions

3. RoBERTa:

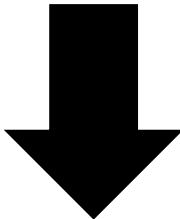
- Natural language approach using a deep-learning architecture fine-tuned on the task of ideological prediction

MNIR

Text processing

- "Bag-of-words" approach

"I do not approve of death taxes"



(‘approv**’, ‘**death**’) (‘**death**’, ‘**tax**’)**

Model specification

- "Bag-of-words" approach
- Select 10,000 most partisan bigrams according to Chi-Squared test
- Fit multinomial model of speech

$$c_i \sim \text{MN} (m_i, q_i(X_i))$$
$$q_{ij}(X_i) = \frac{\exp\left(\alpha_j + \sum_{l=1}^k \varphi_{jl} X_{il}\right)}{\sum_{j \in \mathcal{J}} \exp\left(\alpha_j + \sum_{l=1}^k \varphi_{jl} X_{il}\right)}$$

- Estimate forward regression to obtain model from dimension-reduced word counts to ideology

$$y_i = \beta_0 + \sum_{l=1}^k \beta_l Z_{il}$$
$$Z_{il} = \varphi \frac{c_i}{m_i}$$

MFD

Procedure

- Theoretically validated dictionary from Moral Foundations Theory framework of moral values
- Calculate the relative frequency with which candidates invoke universalist vs. communal rhetoric based on these moral values keywords

$$u = \frac{f_{\text{Ingroup}} + f_{\text{Authority}} - f_{\text{Care}} - f_{\text{Fairness}}}{\text{Total number of non-stop words}}$$

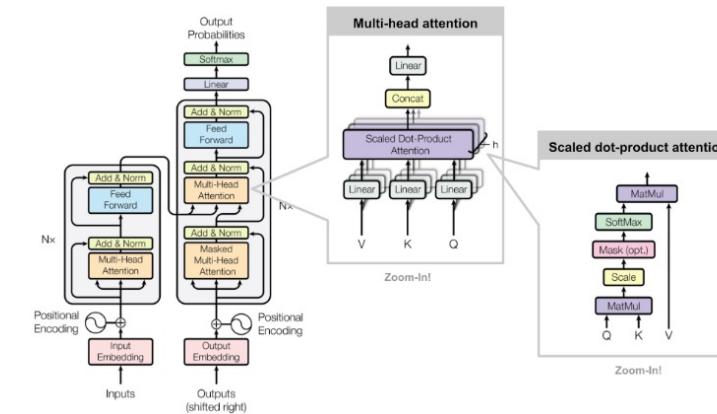
- Universalist rhetoric associated with Democratic party and voting behavior
- Normalize these scores, winsorize, and scale to [-1, 1] to match DW-Nominate

RoBERTa

Model architecture

- State-of-the-art natural language model trained on approximately 160 GB of text data
 - Instead of selecting relevant keywords, this model takes tweets as input, and considers sentence sentiment and grammatical structure

- Transformer architecture with attention mechanisms to "remember" previous words/phrases



- I add a regression head to the architecture and fine-tune the model to the ideological prediction task
 - This updates both the final regression layer coefficient vector as well as the existing model weights

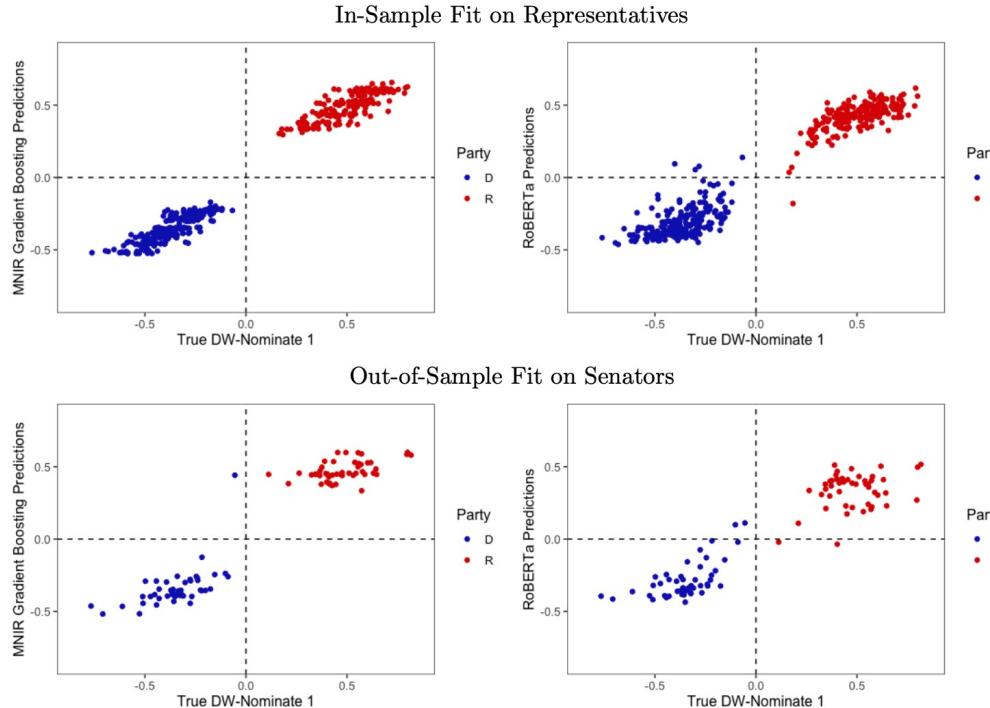
Validation

Assessing the quantitative and qualitative performance of the obtained predictions

- Do these estimates capture ideologically meaningful information?
- How do these estimates perform on and generalize to out-of-sample candidates?

Compelling results quantitatively ...

MNIR, RoBERTa baseline and candidate fit on the 116th and 117th Congress



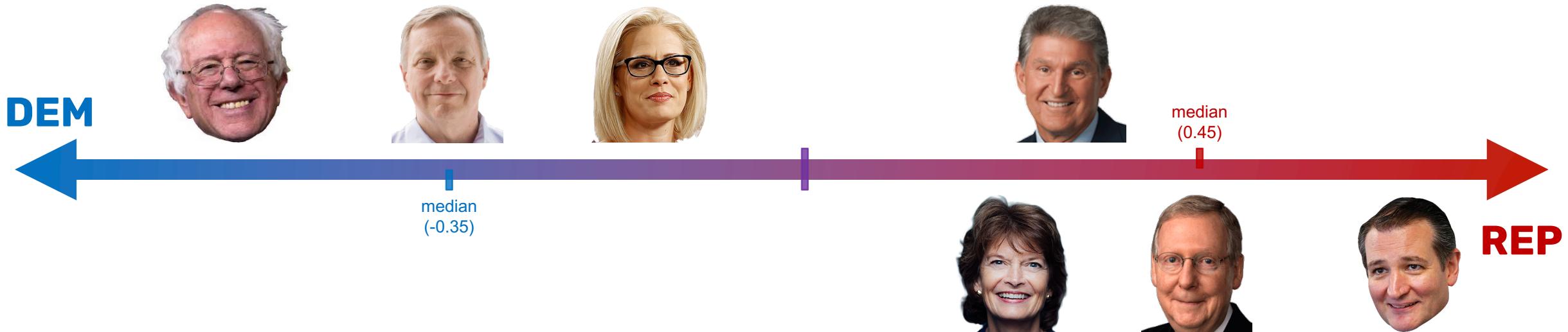
Model	MNIR		RoBERTa
	Gradient Boosting	Regression Forest	
Constant	-0.004 (0.009)	-0.015 (0.008)	0.010 (0.010)
Predictions	1.053*** (0.021)	1.161*** (0.022)	1.434*** (0.035)
N	326	326	326
R ²	0.883	0.896	0.840

Correlation: 0.95 and 0.88 for MNIR, RobERTa

Correlation: 0.99 and 0.96 for MNIR; 0.97 and 0.93 for RoBERTa

... and qualitatively:

Selected senators from MNIR predictions



... and qualitatively:

Most partisan bigrams, keywords from MNIR and MFD

- Identified bigrams, keywords are credible and consistent with traditional policy aims and rhetoric of the two major parties

MNIR

Rank	Baseline Sample		Candidate Sample	
	Most Democratic	Most Republican	Most Democratic	Most Republican
1	gun_violence	tax_reform	health_care	president_realdonaldtrump
2	trump_administration	potus_realdonaldtrump	climate_change	nancy_pelosi
3	climate_change	president_realdonaldtrump	gun_violence	god_bless
4	health_care	speaker_pelosi	working_families	president_trump
5	pre_existing	adam_schiff	public_health	law_enforcement
6	background_checks	pro_growth	mitch_mcconnell	far_left
7	existing_conditions	great_news	voting_rights	radical_left
8	trump_admin	secure_border	affordable_care	thank_realdonaldtrump
9	voting_rights	born_alive	donald_trump	democrat_party
10	#forthepeople_pic	southern_border	social_security	men_women

MFD

Rank	Most Democratic		Most Republican	
	Keyword	Foundation	Keyword	Foundation
1	care	Care Virtue	riot+	Authority Vice
2	equal+	Fairness Virtue	communis+	Ingroup Virtue
3	law	Authority Virtue	patriot+	Ingroup Virtue
4	fight+	Care Vice	order+	Authority Virtue
5	communit+	Ingroup Virtue	illegal+	Authority Vice
6	justice	Fairness Virtue	destroy	Care Vice
7	rights	Fairness Virtue	terroris+	Ingroup Vice
8	families	Ingroup Virtue	legal+	Authority Virtue
9	discriminat+	Fairness Vice	lawless+	Authority Vice
10	defen+	Care Virtue	caste+	Authority Vice

Results

Analyzing the evolution of ideological rhetoric over the election cycle

- Do candidates moderate over the election cycle in accord with the post-primary moderation hypothesis?
- How do candidate- and race-specific characteristics heterogeneously impact this ideological movement?

Empirical Specifications

- Event Study

$$y = X\theta^T + \sum_{t=0}^T \beta_{R,t} \text{Period } t \times \text{Republican} + \sum_{t=0}^T \beta_{D,t} \text{Period } t \times \text{Democrat}$$

- Difference-in-difference with binary general election indicator

$$y = \text{Intercept} + X\theta^T + \alpha \text{Republican} + \beta_{R,G} \text{General} \times \text{Republican} + \beta_{D,G} \text{General} \times \text{Democrat}$$

- Triple difference-in-difference for a binary covariate v (e.g. incumbency status)

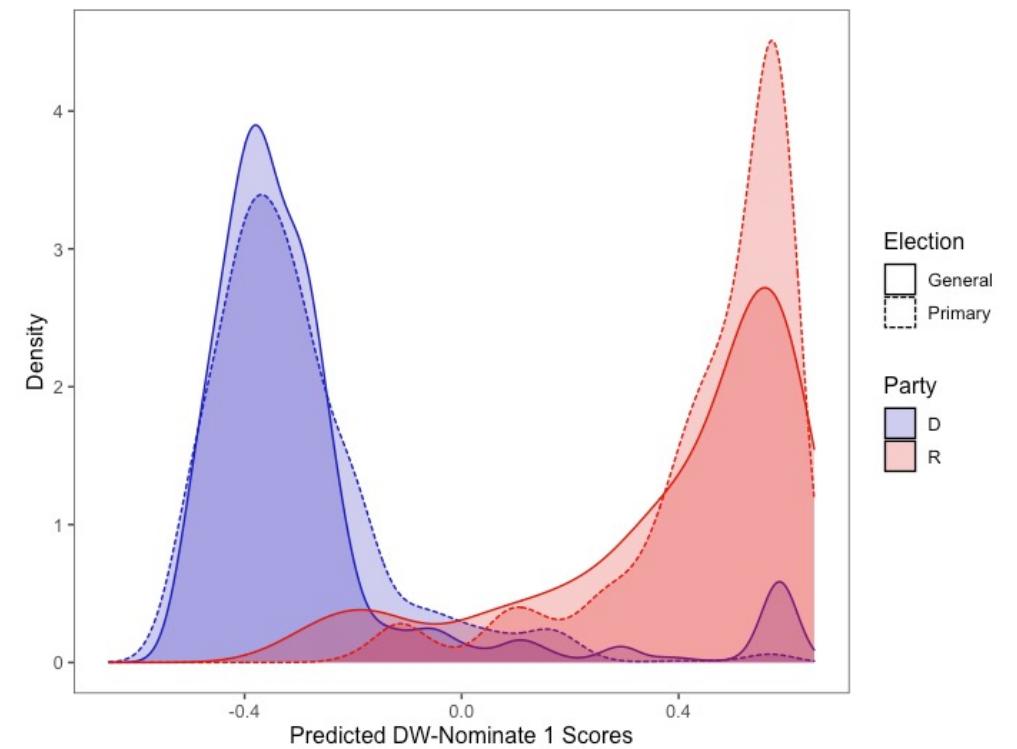
$$\begin{aligned} y = & \text{Intercept} + X\theta^T + (v \times X)\varphi^T + \alpha R + \gamma R \times v + \\ & \beta_{R,v,G} \text{General} \times \text{Republican} \times v + \beta_{R,1-v,G} \text{General} \times \text{Republican} \times (1-v) + \\ & \beta_{D,v,G} \text{General} \times \text{Democrat} \times v + \beta_{D,1-v,G} \text{General} \times \text{Democrat} \times (1-v) \end{aligned}$$

Results

Main Specification

- Across all three methodologies ...
 - Republicans systematically moderate; approximately by half of a standard deviation in the congressional DW-Nominate distribution
 - No effect observed among Democrats
 - In magnitude, Republicans are more extreme over entire cycle than Democrats

	MNIR		RoBERTa		MFD	
	(1)	(2)	(3)	(4)	(5)	(6)
Republican × General	-0.057** (0.022)	-0.058** (0.021)	-0.067*** (0.012)	-0.068*** (0.012)	-0.096*** (0.019)	-0.099*** (0.019)
Democrat × General	0.015 (0.038)	0.014 (0.036)	-0.029* (0.013)	-0.029* (0.013)	-0.024 (0.014)	-0.029* (0.014)
Observations	8,349		8,304		8,349	
R ²	0.609	0.612	0.576	0.593	0.096	0.107
Hypothesis Tests						
$\beta_{R,G} + \beta_{D,G} = 0$	-0.042 (0.045)	-0.044 (0.042)	-0.096*** (0.018)	-0.097*** (0.017)	-0.12*** (0.024)	-0.128*** (0.023)

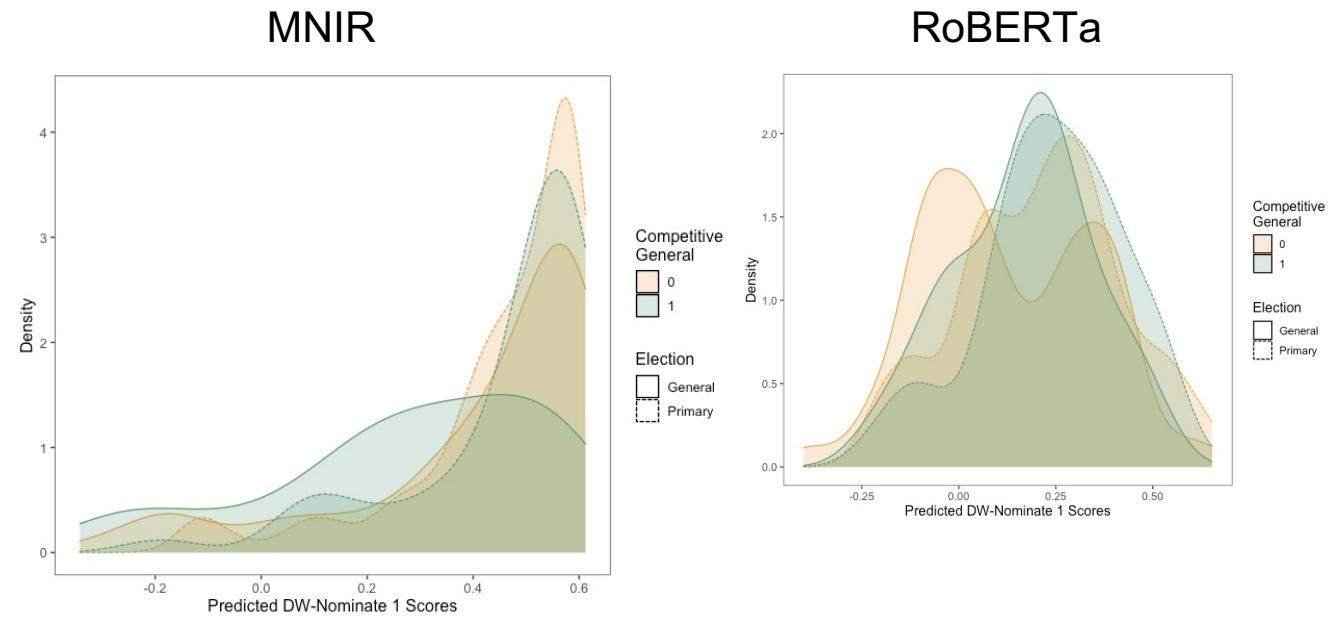


Results

Heterogeneity Specifications

- Mixed evidence that Republicans moderate more in competitive general elections
- No significant difference by incumbency status, primary election competitiveness

	MNIR (1)	RoBERTa (2)	MFD (3)
Republican × Competitive × General	-0.169*** (0.035)	-0.073** (0.023)	-0.108** (0.037)
Republican × Non-Competitive × General	-0.031 (0.025)	-0.067*** (0.013)	-0.095*** (0.021)
Democrat × Competitive × General	0.002 (0.027)	-0.066*** (0.016)	-0.078 (0.043)
Democrat × Non-Competitive × General	0.019 (0.043)	-0.025 (0.014)	-0.017 (0.015)
Candidates in Competitive Races	76	76	76
R ²	0.612	0.579	0.097



Conclusions

- All approaches show compelling, ideologically meaningful quantitative and qualitative validation results
 - Generalizable methodology for obtaining ideology estimates for all candidates
 - Performance should improve as social media becomes more important for political communication
- All approaches show evidence of moderation over the course of the primary
 - Asymmetry in party base extremity may disincentive Democrats to moderate
 - Support for this hypothesis given divergence in overall extremity magnitudes

Thank you for listening!
Questions?



Appendix

Supplemental Tables & Figures

- Sample Summary Statistics
- Event Study Plots
- Main Specification Table
- Heterogeneous Specification Tables
- RoBERTa Relevancy Scores

Sample Summary Statistics

Candidate- and district-specific covariates

	Mean	SD	Median	N
Candidate-Specific				
Incumbent	0.474	0.5	0	665
Republican	0.439	0.497	0	665
Primary Election Competitive	0.068	0.252	0	665
Primary Election Margin	59.8	36.9	64.0	665
District-Specific				
Competitive (Cook PVI)	0.173	0.378	0	423
General Election Competitive	0.097	0.296	0	423
General Election Margin	28.1	20.8	23.7	423
General Election Length (Months)	5.26	2.06	5	423
Trump 2020 Vote Share	0.468	0.152	0.483	423

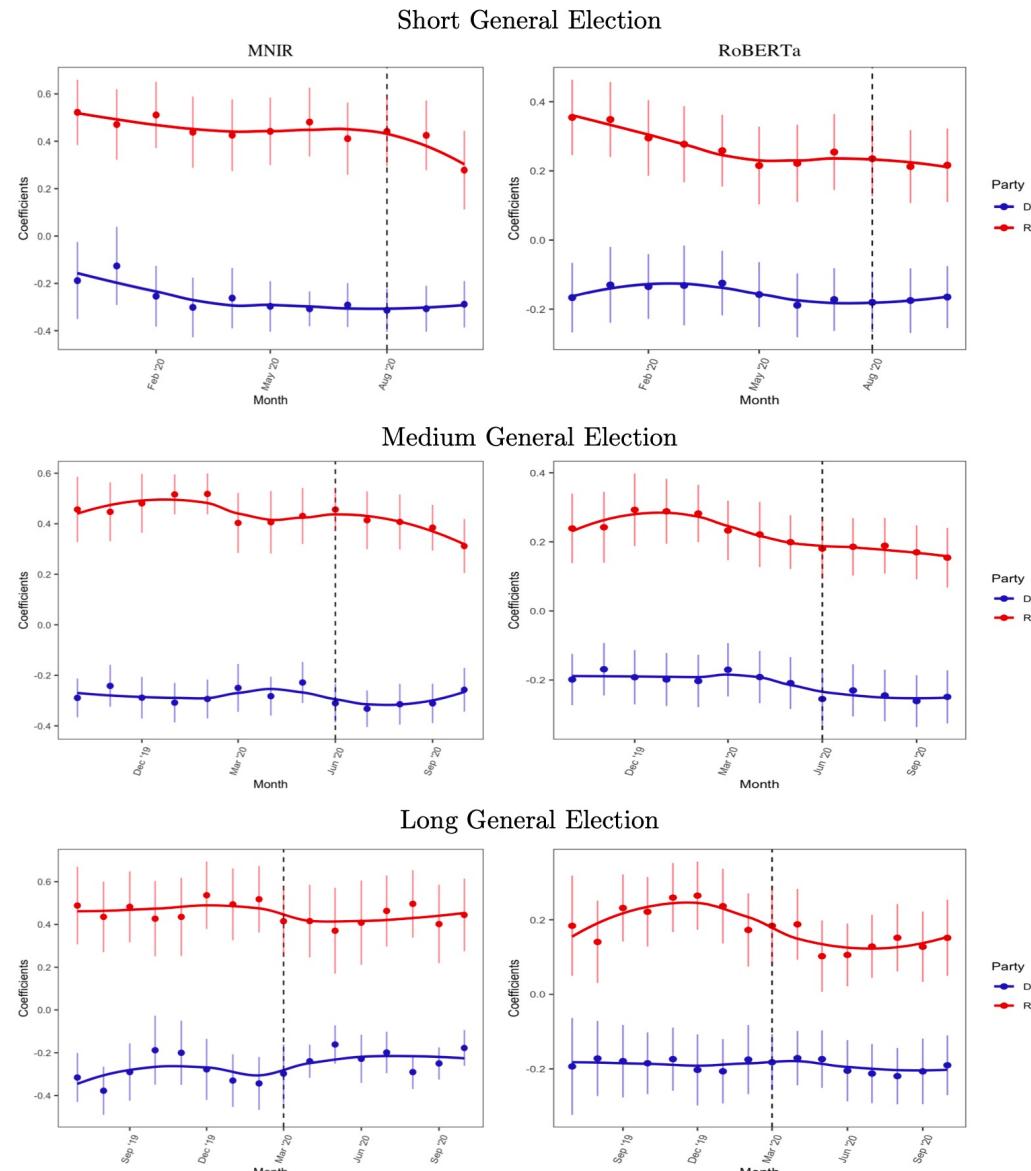
Sample Summary Statistics

Predictions and text counts, by party

	Democrat			Republican		
	Mean	Std. Dev.	Median	Mean	Std. Dev.	Median
Predictions						
MNIR	-0.188	0.361	-0.347	0.396	0.306	0.539
RoBERTa	-0.176	0.186	-0.212	0.216	0.223	0.251
MFD	-0.163	0.349	-0.168	0.051	0.366	0.022
Counts						
MNIR Bigrams	92	122	56	59.6	80.9	36
MFD Care	34.8	46.8	20	18.4	27.5	9
MFD Fairness	8.02	13.4	4	3.08	6.67	1
MFD Ingroup	27.4	36.4	16	17.4	26.3	9
MFD Authority	18.7	26.8	11	16	28	7
Tweet Length (Characters)	13,367	20,026	7,483	8,878	13,950	4,411

Event Study Results

- Notes:
 - Short: 3 month general (19% candidates)
 - Medium: 5 (27%)
 - Long: 8 (25%)



Main Specification

TABLE 8—IDEOLOGICAL MODERATION AMONG 2020 CONGRESSIONAL CANDIDATES

	MNIR		RoBERTa		MFD	
	(1)	(2)	(3)	(4)	(5)	(6)
Republican × General	-0.057** (0.022)	-0.058** (0.021)	-0.067*** (0.012)	-0.068*** (0.012)	-0.096*** (0.019)	-0.099*** (0.019)
Democrat × General	0.015 (0.038)	0.014 (0.036)	-0.029* (0.013)	-0.029* (0.013)	-0.024 (0.014)	-0.029* (0.014)
Republican	0.740*** (0.019)	0.745*** (0.019)	0.410*** (0.019)	0.423*** (0.018)	0.269*** (0.019)	0.268*** (0.019)
Trump 2020 Vote Share	0.339*** (0.082)	0.336*** (0.076)	0.583*** (0.065)	0.618*** (0.070)	0.155* (0.062)	0.147* (0.062)
Incumbent		-0.012 (0.021)		0.039* (0.017)		-0.059*** (0.017)
Senate			-0.025 (0.024)	-0.019 (0.026)		-0.112*** (0.025)
Competitive			-0.044* (0.021)	-0.070*** (0.017)		-0.029 (0.018)
Constant	-0.455*** (0.037)	-0.433*** (0.032)	-0.473*** (0.033)	-0.487*** (0.041)	-0.156*** (0.033)	-0.102** (0.036)
Observations	8,349		8,304		8,349	
Candidates	665		661		665	
Outcome Mean	-0.056		-0.074		0.007	
Outcome SD	0.443		0.274		0.421	
R ²	0.609		0.612		0.576	
Hypothesis Tests						
$\beta_{R,G} + \beta_{D,G} = 0$	-0.042 (0.045)	-0.044 (0.042)	-0.096*** (0.018)	-0.097*** (0.017)	-0.12*** (0.024)	-0.128*** (0.023)

Notes: This table presents the results from estimating Equation 2 on the final candidate sample for the different methodologies. The dependent variable is the predicted DW-Nominate scores from the MNIR model for Columns (1) and (2); the predicted DW-Nominate scores from the RoBERTa model for Columns (3) and (4); and the scaled relative frequency of communal rhetoric for Columns (5) and (6). Odd-numbered columns control only for Trump's 2020 presidential vote share in the district; even-numbered columns also include indicators for incumbency status, competitiveness of the district (Cook PVI), and the congressional chamber. Observations are weighted by the number of bigram counts for the MNIR results, the length of the tweet for RoBERTa, and the number of keyword hits for MFD. All standard errors are clustered at the candidate level. In addition, the results of the hypothesis test on the equality of coefficients for the Republican and Democrat interaction terms are reported.

* Significant at the 5% level.

** Significant at the 1% level.

*** Significant at the 0.1% level.

Competitive General Specification

TABLE 10—IDEOLOGICAL MODERATION, BY GENERAL ELECTION COMPETITIVENESS

	Cook PVI			General Margin		
	MNIR (1)	RoBERTa (2)	MFD (3)	MNIR (4)	RoBERTa (5)	MFD (6)
Republican × Competitive × General	-0.134** (0.042)	-0.088*** (0.017)	-0.117*** (0.029)	-0.169*** (0.035)	-0.073** (0.023)	-0.108** (0.037)
Republican × Non-Competitive × General	-0.025 (0.021)	-0.063*** (0.015)	-0.089*** (0.023)	-0.031 (0.025)	-0.067*** (0.013)	-0.095*** (0.021)
Democrat × Competitive × General	-0.007 (0.016)	-0.056*** (0.015)	-0.081* (0.037)	0.002 (0.027)	-0.066*** (0.016)	-0.078 (0.043)
Democrat × Non-Competitive × General	0.021 (0.048)	-0.024 (0.015)	-0.013 (0.015)	0.019 (0.043)	-0.025 (0.014)	-0.017 (0.015)
Republican	0.761*** (0.022)	0.430*** (0.023)	0.254*** (0.022)	0.736*** (0.021)	0.408*** (0.022)	0.262*** (0.021)
Competitive	-0.225 (0.188)	-0.284 (0.166)	-0.126 (0.259)	-0.793** (0.298)	-0.313 (0.306)	-0.155 (0.397)
Republican × Competitive	-0.065 (0.043)	-0.045 (0.037)	0.064 (0.046)	0.029 (0.041)	0.013 (0.042)	0.055 (0.057)
Trump 2020 Vote Share	0.323*** (0.083)	0.575*** (0.064)	0.164** (0.062)	0.338*** (0.083)	0.583*** (0.065)	0.164** (0.062)
Competitive × Trump 2020 Vote Share	0.476 (0.409)	0.528 (0.352)	0.199 (0.552)	1.583** (0.613)	0.649 (0.620)	0.226 (0.836)
Constant	-0.447*** (0.038)	-0.463*** (0.033)	-0.153*** (0.034)	-0.453*** (0.037)	-0.473*** (0.033)	-0.155*** (0.033)
Observations	8,349	147,666	8,349	8,349	147,235	8,349
Candidates	665	661	665	665	661	665
Candidates in Competitive Races	136	135	136	76	76	76
R ²	0.615	0.59	0.098	0.612	0.579	0.097
Hypothesis Tests						
$\beta_{R,C,G} - \beta_{R,NC,G} = 0$	-0.110* (0.047)	-0.026 (0.023)	-0.027 (0.037)	-0.138** (0.043)	-0.006 (0.027)	-0.013 (0.042)
$\beta_{D,C,G} - \beta_{D,NC,G} = 0$	-0.027 (0.051)	-0.033 (0.021)	-0.068 (0.040)	-0.017 (0.050)	-0.042 (0.022)	-0.060 (0.046)

Notes: This table presents the results from estimating Equation 3 with an indicator for general election competitiveness on the final candidate sample for the different methodologies. For Columns (1)-(3), this indicator identifies candidates in districts with a Cook PVI rating within a four point radius from “EVEN”; and for Columns (4)-(6), candidates in districts with a final general election margin within five percentage points. The dependent variable is the predicted DW-Nominate scores from the MNIR model for Columns (1) and (4); the predicted DW-Nominate scores from the RoBERTa model for Columns (2) and (5); and the scaled relative frequency of communal rhetoric for Columns (3) and (6). All specifications control for Trump’s 2020 presidential vote share in the district. Observations are weighted by the number of bigram counts for the MNIR results, the length of the tweet for RoBERTa, and the number of keyword hits for MFD. All standard errors are clustered at the candidate level. The results of hypothesis tests on the equality of the coefficients for Republican and for Democratic candidates in the general election from competitive vs. non-competitive generals are reported.

* Significant at the 5% level.

** Significant at the 1% level.

*** Significant at the 0.1% level.

Incumbent & Competitive Primary Specification

TABLE 9—IDEOLOGICAL MODERATION, BY INCUMBENCY STATUS

	MNIR	RoBERTa	MFD
	(1)	(2)	(3)
Republican × Incumbent × General	-0.058** (0.018)	-0.050*** (0.013)	-0.134*** (0.032)
Republican × Non-Incumbent × General	-0.058 (0.032)	-0.077*** (0.016)	-0.066** (0.021)
Democrat × Incumbent × General	-0.009 (0.013)	-0.022** (0.008)	-0.008 (0.021)
Democrat × Non-Incumbent × General	0.031 (0.064)	-0.031 (0.020)	-0.047** (0.018)
Republican	0.736*** (0.026)	0.374*** (0.034)	0.265*** (0.025)
Incumbent	-0.021 (0.062)	0.068 (0.080)	-0.059 (0.076)
Republican × Incumbent	0.009 (0.035)	0.168*** (0.046)	-0.009 (0.051)
Trump 2020 Vote Share	0.293** (0.098)	0.552*** (0.124)	0.163 (0.088)
Incumbent × Trump 2020 Vote Share	0.034 (0.125)	-0.236 (0.145)	0.017 (0.164)
Constant	-0.430*** (0.050)	-0.454*** (0.074)	-0.130* (0.054)
Observations	8,349	8,304	8,349
Candidates	665	661	665
Incumbents	315	313	315
R ²	0.609	0.591	0.102
Hypothesis Tests			
$\beta_{R,I,G} - \beta_{R,NI,G} = 0$	0.00 (0.037)	0.028 (0.020)	-0.068 (0.039)
$\beta_{D,I,G} - \beta_{D,NI,G} = 0$	-0.04 (0.065)	0.009 (0.022)	0.039 (0.028)

TABLE 11—IDEOLOGICAL MODERATION, BY PRIMARY ELECTION COMPETITIVENESS

	MNIR	RoBERTa	MFD
	(1)	(2)	(3)
Republican × Competitive Pri × General	0.033 (0.061)	-0.040* (0.019)	-0.118* (0.059)
Republican × Non-Competitive Pri × General	-0.065** (0.022)	-0.070*** (0.013)	-0.095*** (0.020)
Democrat × Competitive Pri × General	0.014 (0.035)	-0.058 (0.035)	-0.040 (0.038)
Democrat × Non-Competitive Pri × General	0.014 (0.039)	-0.027 (0.014)	-0.023 (0.015)
Republican	0.741*** (0.020)	0.411*** (0.021)	0.264*** (0.020)
Competitive Primary	0.119 (0.104)	0.109 (0.062)	0.098 (0.098)
Republican × Competitive Primary	-0.017 (0.070)	-0.043 (0.049)	0.049 (0.074)
Trump 2020 Vote Share	0.365*** (0.084)	0.598*** (0.070)	0.170** (0.064)
Competitive Primary × Trump 2020 Vote Share	-0.318 (0.217)	-0.170 (0.120)	-0.168 (0.196)
Constant	-0.466*** (0.038)	-0.482*** (0.036)	-0.164*** (0.034)
Observations	8,323	147,117	8,323
Candidates	665	661	665
Candidates in Competitive Primaries	46	46	46
R ²	0.612	0.576	0.096
Hypothesis Tests			
$\beta_{R,CP,G} - \beta_{R,NCP,G} = 0$	0.098 (0.065)	0.029 (0.023)	-0.023 (0.062)
$\beta_{D,CP,G} - \beta_{D,NCP,G} = 0$	0.00 (0.052)	-0.031 (0.038)	-0.017 (0.041)

RoBERTa Relevancy Scores

Word Importance

#s C for years , don ald trump paid zero in federal income taxes . he paid only \$ 750 in the year of his election and the first year of his presidency : trump continues to use government for his own benefit . it 's our job to stop him . we must vote for jo eb iden . how it 's done . you know who supports a peaceful transfer of power ? these two . tune in tonight at pm et to watch jo eb iden take on trump in the first presidential debate . bid en har ris 2020 you know who releases their tax returns ?, , thank you jo eb iden for exposing trump 's fantasy to take away health care from tens of million of americ ans . if trump wins reelection , his fantasy is likely to become reality ., here st hed eal : when trump has nothing left , he gr as ps at his infant ile slurs . we need jo eb iden ., trump has no plan ., thank you jo eb iden for explaining the timeline of trump 's reckless , dangerous incompetence on fighting the coron av irus . we need a self less leader . we need jo e bid en ., pres . trump lied to the american people about coron av irus and how deadly it is . C #/s

Word Importance

#s C trump war room real donald trump pars cale americ ans has the duty to vote dem s out . this is insane . playing with americ ans life is totalitarian ism . us a will never be a socialist country . president trump americ ans want four more years ., real donald trump americ ans have the duty to vote dem s out . this is insane . playing with americ ans life is totalitarian ism . us a will never be a socialist country . president trump , americ ans want four more years ., ch ris j z ull o real donald trump , trump war room lat am fr press sec sc av ino 45 i am an american , a republican running for us congress against communist democrat coup inst igator to make ge org iac d 4 great again with trump vict ory team and black vo ices fort r ump we will re - elect pres trump C #/s