Campaign Dynamics, Primary Elections, and Extremism: Do U.S. Congressional Candidates Moderate Their Message After Winning Their Primary?

Abstract

Though candidates' messaging strategies are broadly characterized as dynamic, scant research directly examines how candidates modify the ideological extremity of their messages throughout the campaign. In this note, I argue that candidates who win their primaries should moderate their communications to target general-election voters in the period immediately following their primary election. To test this, I develop a continuous measure of the expressed ideological position of those candidates based on more than 200,000 Facebook posts from U.S. congressional candidates in the 2016 election. The measurement method extends existing network-based measures of ideology in an on-line framework to greatly increase the number of observations in the scaling process. Using a two-way fixed effects model that approximates a difference-in-differences strategy, I find that winning a competitive primary leads candidates to moderate their expressed ideology, but that this shift is modest relative to the baseline trend towards moderation among candidates without primary competition.

Both scholars and political commentators have long speculated that candidates running for office strategically alter their communications over the course of a campaign to maximize their probability of election. One common theory is that candidates moderate their speech after the primary stage of their election in pursuit of moderate voters (Acree et al. 2020; Burden 2001). However, the theoretical justification for this claim is mixed, and empirical support is scarce.

Theoretically, existing models lead to conflicting expectations about how candidates may change their rhetoric after a primary victory. In one view, post-primary moderation can make candidates more attractive to moderates, but at the potential cost of appearing inconsistent or untrustworthy (Agranov 2016; Hummel 2010). In contrast, standard two-stage models of elections simply assume that candidates will remain ideologically consistent throughout (Owen and Grofman 2006).

Empirically, evidence indicates that legislators condition their public communications on both the preferences of their constituents and their own voting record in Congress (Fenno 1978; Grimmer et al. 2014). This suggest that candidates can weigh the relative costs of appearing inconsistent against the benefits of appealing to voters' preferences, and thus might modify what I call their *expressed ideology* over the course of a campaign. However, with the exception of Acree et al. (2020), which analyzes moderation among only three presidential candidates, existing studies of campaign dynamics have relied on non-rhetorical data such as roll-calls (Burden 2001). In essence, we know very little about the effects of winning a competitive primary election on campaign rhetoric.

The scarcity of research on campaign dynamics is, in part, a consequence of inadequate data. Many congressional races do not run radio or television advertisements, nor do they receive sufficient news coverage to develop a complete and direct measure of candidates' ideological expression. Candidate websites offer a more complete record of candidates' expressed ideology (Druckman et al. 2010), but have historically been relatively

static, and thus may not capture mid-election shifts in ideological expression. Moreover, even when candidates' rhetorical data is available throughout an election, it can be difficult to extract a straightforward unit of measure that applies across candidates and time.

In this note, I test the theory that candidates who win their primary election moderate their expressed ideology using social media posts. This approach has three key advantages. First, prior research finds that candidates use social media to complement their offline political communications (Kreiss et al. 2018; Larsson 2016), suggesting that ideological expression in their posts should mirror conventional communications. Second, because social media posts are connected in a network by users' engagements, their interconnectedness allows them to be placed on a meaningful unidimensional scale that reflects their ideological content (Bond and Messing 2015). This is advantageous because it allows for a combination of text and audio-visual communications to be simultaneously scaled in the same measure. Of the posts used in this analysis, roughly one-third are classified as image or video content, which would be excluded using conventional text methods.

Third, social media posts provide fine-grained data over time. Among candidates analyzed below, House candidates produced more than one post per week and Senate candidates averaged a post every four weeks (details in Appendix B). Having numerous observed updates to a candidate's expressed ideology on a weekly basis offers better coverage than the comparatively-infrequent updates to print, radio, television, or websites.

Leveraging the benefits of social media data, I investigate whether candidates in the 2016 U.S. congressional elections implement dynamic messaging strategies to express different ideological positions before and after winning their primary election. Using a novel dataset of Facebook posts authored by these candidates, I construct a weekly measure of candidates' expressed ideological extremity and use it to test the hypothesis that candidates engage in post-primary moderation, whereby their expressed ideological

extremity decreases after winning their primary.

To causally identify the effect of primary victory, I compare the rates of moderation among candidates with primary competition against those without competitive primaries to account for unobserved confounding. I accomplish this using a linear two-way fixed effects model, which approximates a multi-period difference-in-differences analysis. I find evidence that candidates in contested primaries do moderate after winning their primary, but that even candidates without primary competition moderate their expressed ideologies throughout the electoral period. In Appendix F.5 I show that these findings are robust to alternate specifications, including additional lagged dependent variables, as well non-parametric and synthetic control time series cross-sectional analyses. This finding contributes to an enhanced understanding of the role that primaries play in ideological polarization and introduces a method by which additional research can be directed towards questions of intra-electoral ideological dynamics.

Data

To test my hypothesis, I study Facebook posts authored by U.S. congressional candidates in the 2016 election. These posts are part of a novel dataset collected through Facebook's Graph API in July 2017 encompassing all posts published on the official, public pages (as defined in Appendix A) of legislators elected to the 115th U.S. Congress. The corpus consists of 840,909 posts authored by 425 House members and 97 Senators from July 2017 to the day on which each legislator's account was created.

To measure the expressed ideology of individual social media posts, I adapt the technique of Bond and Messing (2015), which estimates the ideology of politicians' Facebook pages based on regularities in the network formed by users who endorse each page with a like. Whereas Bond and Messing (2015) applies this method to Facebook pages, I apply it to individual Facebook posts by forming the network which connects users' likes to politicians' posts (see Appendix C.1). I further adapt the method by computing the latent

scores using an observation-wise implementation of principal component analysis (technical details in Appendix C.2), which overcomes the computational costs of performing matrix decomposition on a matrix with billions of cells.

The absolute value of these latent ideology scores measures the overall expressed ideological extremity of each post, regardless of whether the content is liberal or conservative. For analysis, I construct a weekly time series by averaging ideological extremity across candidates and weeks.[1] For weeks in which candidates have no expressed ideology, I use multiple imputation to create 10 datasets which estimate missing entries by conditioning on observed data. Additional details on the imputation method are available in Appendix E. The dependent variable for analysis, weekly expressed ideological extremity, represents the average ideological extremity of a candidate's posts produced in a given week. For all candidates, scores closer to 0 indicate more moderate expressed ideology than large scores.

Intuitively, the network used to estimate these scores is like a human-coded dataset where each of 14,666,834 users is presented with a binary choice to endorse the sentiment of each politician's posts. Assuming that each user responds with their own stable ideological convictions in mind, the aggregate result of these 287,624,351 likes separates posts along ideological lines according to the composition of unique users who agree with a given post. The advantage of this approach is that users make their endorsement decisions based on the combination of the post's content and their prior knowledge about its broader context. This allows the scaling to incorporate a post's literal text, its figurative language, and non-text features such as images, audio-visuals, and linked content.

I implement this procedure on the subset of legislators' Facebook posts which received 100 or more likes as of July 2017. This includes 204,171 unique posts in the scaling process. I extract 27,430 posts ranging from November 11, 2015 to November 15,

[1]Weeks begin Wednesdays and end Tuesdays so elections fall at the end of a week.

(a) -0.97 (b) 0.02 (c) 0.73

Figure 1: Example Facebook Posts with Expressed Ideology Scores

2016 (from 52 weeks before to one week after the 2016 general election) for analysis, incorporating posts authored by 375 congressional candidates (31 Senate; 344 House).[2] Figure 1 contains example posts with their corresponding expressed ideology scores. I conduct extensive validation on the measure in Appendix D. There I present randomly-selected posts and their corresponding scores, the distribution of ideology scores across party lines, a classification exercise using scores to predict the ideology of hand-labeled posts, analyses of within-candidate and within-state variance, the internal consistency of the measure, and evidence that ancillary network features do not affect the measure.

Estimation

The post-primary moderation hypothesis predicts that candidates who win primary elections should decrease their expressed ideological extremity after transitioning into the

[2]This excludes 66 incumbent senators not up for election in 2016, 2 senators and 87 House members who do not produce posts during the 2016 election, as well as 1 additional senator and 4 House members from Louisiana. Appendix B contains additional information about candidates excluded from analysis.

general election. However it is also possible that all candidates (including those without primary competition) might moderate before the general election to court moderate voters. To determine whether primary candidates moderate their ideological expression after their primary, I compare their post-primary communications against candidates who did not face primary competition.[3] The comparison of these two groups of candidates creates a difference-in-differences-like design, where completing a primary contest represents a treatment intervention, candidates in competitive primaries are susceptible to treatment, and candidates without primary competition are not susceptible to treatment. The treatment group (hereafter, the *primary group*) consists of 197 candidates (179 House; 18 Senate) who ran in competitive primaries, while the control group (the *no primary* group) contains 178 candidates (165 House; 13 Senate) who ran in uncompetitive primaries.

Because primaries are staggered over many months, the difference-in-differences framework must be extended as a time series. The conventional method for causal inference with panel data is the linear two-way fixed effects model (TWFE), which compares the difference between treated and controlled units while simultaneously controlling for unobserved confounders affecting individual units and time periods. Though more sophisticated methods have emerged in recent years, I focus on TWFE in the main text due to its intuitive appeal and widespread use. Alternate models using non-parametric fixed effects and synthetic controls are presented as robustness checks in Appendix F.5. Their results support the conclusions derived from the TWFE approach described here.

In addition to assuming that temporal and unit-specific confounding remains constant over time (and hence can be controlled with intercept terms), TWFE requires that treatment be uncorrelated with past outcomes. This assumption is implausible here since

[3]Primaries are coded as uncompetitive if the 2016 Federal Election Commission results indicate that the candidate (a) ran unopposed, (b) was nominated by party convention without a primary runoff, or (c) won their primary contest with 95% or more of the vote.

it requires that a candidate's pre-treatment expressed ideology have no effect on their electoral outcome in the primary. Therefore I estimate a linear TWFE model with a lagged dependent variable to account for the effect of past outcomes, and yield a causal estimate of the effect of winning a contested primary election on expressed ideological extremity. The model is of the form:

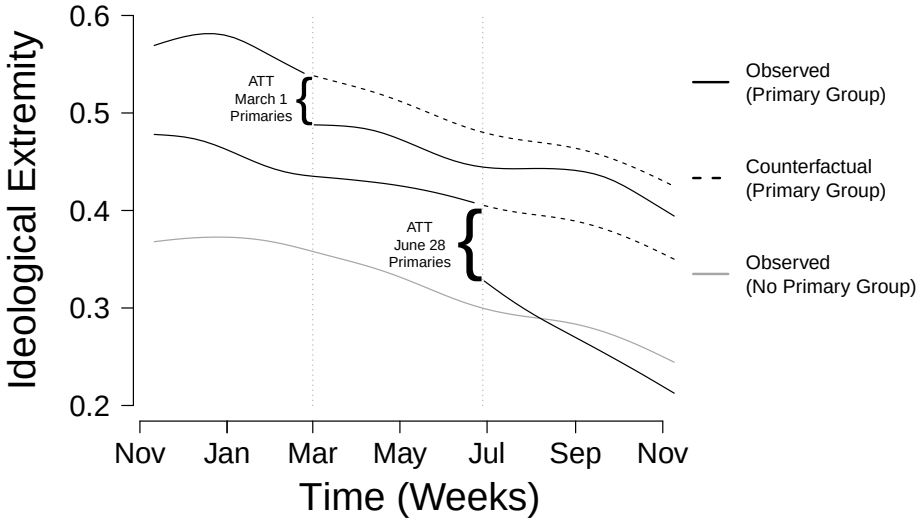
$$y_{it} = \beta D_{it} + \sum_{l=1}^{L} \rho_l y_{i,t-l} + \alpha_i + \omega_t + \epsilon_{it} \tag{1}$$

for candidate i in week t, and where y_{it} is the outcome (expressed ideological extremity), β is the effect of treatment, D_{it} is a treatment indicator, ϵ_{it} is residual error, and α_i and ω_t represent candidate and week intercepts, respectively. ρ_l is the coefficient on the lagged dependent variable, $y_{i,t-l}$, with corresponding lag l. The model presented below uses a one-period lag, but I demonstrate that the results are robust to alternate lag specifications in Appendix F.4. I estimate the model with panel-corrected standard errors to account for potential dependencies in the errors imposed by candidates and time.

Results

To illustrate the effects of post-primary moderation, Figure 2 presents the mean trends (averaged across 10 imputed datasets and smoothed by cubic splines) in expressed ideological extremity for all candidates in the *no primary* group (grey) and two cohorts of candidates in the *primary* group (black) who participated in primaries on either March 01, 2016 or June 28, 2016 (Appendix F.1 contains trend lines for all primary cohorts). For each primary cohort, I estimate pre- and post-primary splines (solid lines) and plot those along with the post-primary counterfactual trend line as measured by the *no primary* group's average. The trend lines indicate that in the year preceding the 2016 general election, both candidates in competitive and uncompetitive primaries moderated the ideological extremity of their Facebook posts as the general election approached. The post-primary

Figure 2: Average Expressed Ideology over Time (by Primary Cohort)



Average trends in expressed ideological extremity, smoothed by cubic spline, for all candidate in uncompetitive primaries (grey) and two primary cohorts in competitive primaries (black). Solid lines indicate observed values and dashed lines indicate counterfactual values parallel to the *no primary* group. Appendix F.1 contains trends for all primary cohorts. The data ranges from November 11, 2015 to November 15, 2016, and consists of average values in expressed ideological extremity across 10 multiply-imputed datasets.

trend lines indicate that those who survive competitive primary contests produce less ideologically extreme posts in their general election phase than one would predict given the average amount of moderation exhibited by the *no primary* group.

The TWFE model presented in Table 1 estimates the average treatment effect among candidates in competitive primary elections (ATT). It supports the post-primary moderation hypothesis. The coefficient on treatment is negative and statistically distinguishable from 0 (coef: -0.021, p: 0.000), indicating that candidates who won competitive primaries reduced their expressed ideological extremity more than candidates without primary competition in the week after the primary occurred. However, this moderation is equivalent to only 4% of a standard deviation in the dependent variable. By comparison, the average difference in expressed ideology between the beginning and end of the time series is 0.128 and 0.123 for the *primary* and *no primary* groups, respectively. These dif-

ferences are approximately one-third of a standard deviation in the dependent variable. Thus the effect of post-primary moderation is meager relative to the overall degree of moderation that all candidates exhibit during the entire electoral period.

Table 1: Two-way Fixed Effects Results

Treatment	**-0.021**
	(0.005)
Candidate & Week FEs	✓
N Candidates	375
N Observations	19,875

Two-way fixed effects regression results. Panel-corrected standard errors in parentheses. Coefficients with bold text are statistically significant at the 0.05 level.

Conclusion

Though past research treats ideological expression as a static choice, evidence in this note suggests that ideological expression is a dynamic process. Using congressional candidates' Facebook posts I find the most fine-grained evidence to date of post-primary moderation among congressional candidates. The statistical analysis reveals that candidates in competitive primary elections moderate their expressed ideology immediately following primary victory, but the substantial moderation among candidates without primary competition suggests that all candidates express themselves more moderately in advance of the general election. These moderating trends in ideological expression suggest a more complex relationship between primary elections and congressional polarization than previously theorized, however this note provides the methodological tools by which future research can connect ideological campaign rhetoric with its ultimate effect on polarized congressional representation.

References

Acree, Brice DL et al. (2020). "Etch-a-sketching: Evaluating the post-primary rhetorical moderation hypothesis". *American Politics Research* 48.1, pp. 99–131.

Agranov, Marina (2016). "Flip-flopping, Primary Visibility, and the Selection of Candidates". *American Economic Journal: Microeconomics* 8.2, pp. 61–85.

Bond, Robert and Solomon Messing (2015). ""Quantifying social media's political space: Estimating ideology from publicly revealed preferences on Facebook"". *American Political Science Review* 109.1, pp. 62–78.

Burden, Barry C (2001). "The polarizing effects of congressional primaries". *Congressional primaries and the politics of representation*, pp. 95–115.

Druckman, James N et al. (2010). "Issue Engagement on Congressional Candidate Websites, 2002 - 2006". *Social Science Computer Review* 28.1, pp. 3–23.

Fenno, Richard Francis (1978). *Home style*. Little & Brown.

Grimmer, Justin, Sean J Westwood, and Solomon Messing (2014). *The impression of influence: legislator communication, representation, and democratic accountability*. Princeton University Press.

Hummel, Patrick (2010). "Flip-flopping from primaries to general elections". *Journal of Public Economics* 94.11-12, pp. 1020–1027.

Kreiss, Daniel, Regina G Lawrence, and Shannon C McGregor (2018). "In their own words: Political practitioner accounts of candidates, audiences, affordances, genres, and timing in strategic social media use". *Political communication* 35.1, pp. 8–31.

Larsson, Anders Olof (2016). "Online, all the time? A quantitative assessment of the permanent campaign on Facebook". *New Media & Society* 18.2, pp. 274–292.

Owen, Guillermo and Bernard Grofman (2006). "Two-stage electoral competition in two-party contests: persistent divergence of party positions". *Social Choice and Welfare* 26.3, pp. 547–569.

Campaign Dynamics, Primaries, and Extremism: Do U.S. Congressional Candidates Moderate After Winning Their Primary?

Supplementary Appendix

Table of Contents

A Terminology **A-3**

B Data Description **A-4**

C Measurement **A-7**

C.1 Measuring Ideology from Social Media Data A-7

C.2 Online Dictionary Learning A-11

D Validation **A-15**

D.1 Examples of Expressed Ideology A-16

D.2 The Distribution of Expressed Ideology A-19

D.3 Within-Candidate Variation A-23

D.4 Within-State Variation A-24

D.5 Temporal Stability A-27

D.6 Changing Rhetoric or a Changing Network? A-28

E Imputation **A-33**

F Robustness **A-36**

F.1 Trends in Expressed Ideological Extremity by Primary Cohort A-37

F.2 Balance Test for Primary and No Primary Groups A-39

F.3 Testing for Non-Stationarity and Unit Root A-41

F.4 Alternate Lag Specifications A-42

F.5 Alternate Estimators A-43

G Appendix References **A-46**

The brevity of the short article format requires saving technical details for the appendix. This appendix includes six main sections. Appendix A operationalizes social-media-related terminology used throughout the main text and appendix. Appendix B contains descriptive statistics about legislators included and excluded from analysis, as well as details about the various primary elections that occur within the analysis and the number of legislators involved in each. In Appendix C I provide technical details on estimating ideology scores from social media data (Appendix C.1) and data reduction with online dictionary learning (Appendix C.2). To validate expressed ideology scores I provide extensive description of the scores in Appendix D. This includes a demonstration of randomly-selected example posts (Appendix D.1), distributional evidence and a classification exercise (Appendix D.2), within-candidate analysis of scores (Appendix D.3), within-state analysis of scores (Appendix D.4), evidence of the scores' consistency across years (Appendix D.5), and diagnostics to rule out the effect of non-rhetorical post features on the measurement of expressed ideology (Appendix D.6). In Appendix E I describe the multiple imputation method used to fill missing expressed ideology scores. Finally, I conduct robustness checks on the analytical result in Appendix F. This includes visual checks for parallel trends between *primary* and *no primary* groups (Appendix F.1), a test of balance between *primary* and *no primary* groups (Appendix F.2), diagnostics for time trends in the dependent variable (Appendix F.3), alternate lag specifications for the two-way fixed effects model (Appendix F.4), as well as alternate time series causal inference estimators (Appendix F.5).

A. TERMINOLOGY

The main text references features of the Facebook social media platform present at the time of writing. This section of the appendix operationalizes terms used throughout the paper.

A Facebook user is defined as an individual's social media account on the Facebook platform. Users can upload ("post") content to Facebook in a variety of media including text, image, video, or links to other websites. Users may post this content onto their public profile ("wall") or interact with other users by posting content in response to others' posts.

A "like" is considered an endorsement of some content by the initiating user, and indicates that the user agrees with or supports the content that they have endorsed (Bond and Messing 2015). Users can apply likes to all content on Facebook, including posts, photos, events, etc. Importantly, the like is distinct from a number of other reactions meant to send different signals about the content. In addition to the like, these reactions consist of emoticons representing "love," "haha," "wow," "sad," and "angry."

A Facebook page is a publicly-accessible account owned by a public figure or organization. Pages post content to disseminate information to users, though users may engage with this content in response. Pages are managed by users, who may have their own personal user accounts which post material that is not (necessarily) visible on the page. The formal distinctions between what pages and users can do on Facebook are few, though one key distinction is that pages can be "verified" by the social media platform. Verified pages are noted by the symbol (a white checkmark circumscribed in blue) placed to the right of a page's name, and which indicates that the page is an official representation of the person or organization for which the page is named.

B. DATA DESCRIPTION

Here I present descriptive statistics for the study's data in this appendix. Table A1 presents the amount of Facebook posts produced by congressional candidates in the year before the 2016 general election. Post totals and averages are shown for candidates overall (columns 1), as well as by party (columns 2 -3) and by chamber (columns 4 - 5). As referenced in the text, the average candidate produces more than one post per week. This holds true for the average Democrat and the average House member, though Republicans and senators tend to author posts at lower rates.

Table A1: Congressional Facebook Posts

	All	Democrats	Republicans	House	Senate
Posts	27,430	12,022	15,408	22,837	4,593
Avg. Posts per Candidate	73.15	71.56	74.43	66.39	148.16
Avg. Daily Posts per Candidate	0.20	0.09	0.11	0.16	0.03
Avg. Weekly Posts per Candidate	1.38	0.60	0.78	1.15	0.23

Facebook posts produced by candidates elected to the 115th Congress and which received more than 100 likes. Observations range from November 11, 2015 to November 15, 2016, and contain posts authored by 31 Senators and 344 House members.

Table A2 details the frequencies of candidates included in analysis by chamber and party (upper section). It also describes candidates excluded from analysis (lower section). Of those excluded from analysis, 66 were senators not up for election in 2016, 89 were candidates without active Facebook activity in the period of study, and 5 were candidates in Louisiana. Notably, all candidates included in analysis won their congressional elections in 2016. This feature of the analysis is due to data availability constraints. By the time of scraping in July 2017, many official pages for defeated congressional challengers had been deactivated, and their contents were rendered inaccessible. Including the few challengers' pages would likely offer only a biased understanding of how congressional challengers behaved during the 2016 election. Instead, focusing solely on candidates who were subsequently elected to Congress provides a complete understanding of electorally-

successful ideological messaging strategies.

Table A2: Number of Candidates Included and Excluded from Analysis

	— Senate —		— House —	
	Democrats	Republicans	Democrat	Republican
Analyzed	10	21	158	186
Excluded	38	31	36	55
Not Running	36	30	0	0
No Posts	2	0	35	52
Louisiana	0	1	1	3

Descriptive statistics for legislators in the 115th Congress excluded from analysis. "Not Running" indicates senators who were not up for election in 2016. "No Posts" indicates candidates who either had not created an official, public account at the time of scraping (July 2017), or who had an account but did not produce posts with more than 100 likes in the period of study. "Louisiana" indicates elected representatives for the state of Louisiana, where primary elections do not occur prior to the general election date.

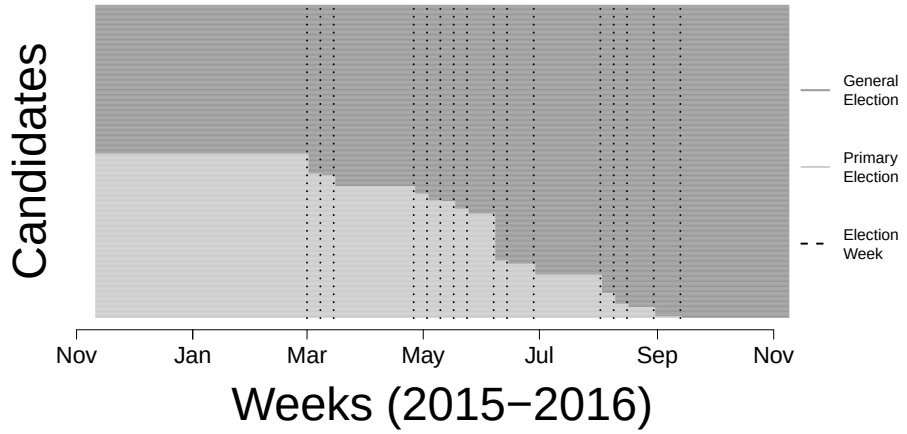
Table A3 describes how candidates in the analysis break down by primary cohort and across *primary* and *no primary* groups and Figure A1 offers a visualization of when candidates enter the general election phase. Each horizontal line corresponds with a unique congressional candidate. Dark lines indicate candidate who are focused on their general election campaign whereas light grey lines indicate candidates who are focused on competitive primary campaigns. Candidates in the *no primary* group are those candidates who are focused on their general elections (dark grey lines) at the beginning of the time series.

Table A3: 2016 Primary Elections and Cohort Sizes

Date	State(s)	Cohort	Primary	No Primary
Mar. 01	AR, AL, TX	41	24	17
Mar. 08	MS	4	2	2
Mar. 15	IL, NC (Senate), OH	32	13	19
Apr. 26	MD, PA	22	9	13
May 03	IN	8	8	0
May 10	NE, WV	5	1	4
May 17	ID, KY, OR	14	9	5
May 24	GA	12	6	6
Jun. 07	CA, IO, MT, NJ, NM, NC (House), SD	73	56	17
Jun. 14	ME, NV, ND, SC, VA	21	4	17
Jun. 28	CO, NY, OK, UT	42	13	29
Aug. 02	KS, MI, MO, WA	31	22	9
Aug. 04	TN	8	5	3
Aug. 09	CT, MN, VT, WI	19	8	11
Aug. 13	HI	2	2	0
Aug. 16	AK, WY	2	2	0
Aug. 30	AZ, FL	26	11	15
Sep. 08	MA	9	0	9
Sep. 13	DE, NH, RI	4	2	2

Chronological listing of 2016 primary election dates by state. Counts indicate the number of candidates included in each primary cohort and how they separate into *primary* and *no primary* groups.

Figure A1: Election Stages for Candidates Over Time



Visualization of candidates entering the general election stage. Each row represents a candidate, whose primary status is indicated by the shading of their cell at week t. The first week begins on November 11, 2015 and the final week ends on November 15, 2016.

C. MEASUREMENT

This section of the appendix details the computation of expressed ideology scores. The first subsection (Appendix C.1) explains the procedure for creating a network matrix formed by users' likes and politicians' posts, as well as the steps required to pre-process the matrix before data reduction. The second subsection (Appendix C.2) explains the method of data reduction, Online Dictionary Learning, which I use to iteratively estimate the principal components of the network matrix.

C.1. *Measuring Ideology from Social Media Data*

As the main text describes, I adapt the ideal point estimation method proposed by Bond and Messing (2015) to measure the affiliation between Facebook posts authored by candidates in the 2016 U.S. congressional elections. Bond and Messing (2015) originally applies the method to political officials of various types by looking at the network of endorsements across pages. It applies singular value decomposition (SVD) to reduce the matrix's data to a unidimensional scale. The measure is validated by comparing the scores to both the first dimension of NOMINATE and to user-reported measures of political ideology. Across both instance, the measure provides consistent estimates of ideology.

I adapt this method by applying it to specific social media posts rather than whole social media pages. To do so, I first form a matrix, $\mathbf{E}^{u \times p}$, with u unique users on the rows, p unique posts on the columns, and cells populated by indicators for whether a user liked a post (1 if true; 0 if false). If a user in this matrix only likes a single post, their row is removed from the matrix, as they can provide no information about the relationship between any two posts. Additionally, I remove the columns of posts which receive fewer than 100 likes from this matrix. Posts with too few likes tend to result in unreliable estimates of expressed ideology. Therefore I implement a floor of 100 likes for all posts in the network matrix in order to ensure reliable ideal point estimates for each post. Future

research should work to determine the exact relationship between the number of likes a post receives and the reliability of its measured ideal point.

The network matrix describes the relationship between users and posts, but does not directly model the relationship between one post and another. Forming the matrix that relates posts to one another requires pre-multiplying the network matrix by its transpose, resulting in a square affiliation matrix, $\mathbf{F}^{p \times p}$, which contains rows and columns equal to the number of unique posts.

$$\mathbf{F} = \mathbf{E}^{\mathbf{T}}\mathbf{E}$$

Rather than modeling the relationship between users and posts (as in **E**), **F** now explicitly models the relationships between each dyad of posts based on the number of users who like both messages. The diagonals of **F** consist of the total number of likes for a given post, and the off-diagonals indicate the number of users who like both the post indicated by the row and the post indicated by the column (see Table A4, middle panel).

Though this affiliation matrix relates posts to one another, the entries of **F** vary widely as a consequence of each post's overall popularity. Unmodified, these differences in popularity weight the ideology scores of posts by their number of likes — pushing popular messages to the fringes of the ideological distribution (Bond and Messing 2015). This requires normalizing each of the rows in **F** to sum to 1 by dividing each row by its diagonal element.

$$\mathbf{G} = \begin{bmatrix} \mathbf{F}_1 / diag(\mathbf{F}_1) \\ \mathbf{F}_2 / diag(\mathbf{F}_2) \\ \dots \\ \mathbf{F}_p / diag(\mathbf{F}_p) \end{bmatrix}$$

Entries in this new agreement matrix, $\mathbf{G}^{p \times p}$, now represent the proportion of all

users who liked the row post and also liked the column post (see Table A4, bottom panel). The values in **G** range from 0 (no common likes) to 1 (all common likes; the diagonal elements), and are irrespective of each post's original popularity. An example matrix undergoing each step in the pre-processing sequence is depicted in Table A4.

Table A4: Pre-processing the User-Post Matrix

Example Network Matrix, **E**

	"Repeal and replace"	"Obamacare works"	...	"It's a human right"
User A	1	0	...	0
User B	0	1	...	1
User C	0	0	...	1
User ...	...	...	...	...
User X	0	0	...	0
User Y	1	0	...	1
User Z	0	1	...	0

Example Affiliation Matrix, **F**

	"Repeal and replace"	"Obamacare works"	...	"It's a human right"
"Repeal ..."	4,784,474	135,603	...	36,201
"Obamacare ..."	135,603	171,686	...	3,456
...	...	...	...	...
"It's ..."	36,201	3,456	...	47,951

Example Agreement Matrix, **G**

	"Repeal and replace"	"Obamacare works"	...	"It's a human right"
"Repeal ..."	1	0.028	...	0.008
"Obamacare ..."	0.790	1	...	0.020
...	...	...	...	...
"It's ..."	0.755	0.072	...	1

Examples of matrices involved in the pre-processing stages of the network-based ideology measure. Matrix **E** relates Facebook users' likes to candidates' posts. Matrix **F** relates candidates' posts to one another by the number of common users who liked both posts indicated by each respective row and column. Matrix **G** relates candidates' posts to one another by the proportion of users who liked the column post among all users who liked the row post.

The agreement matrix contains all of the necessary information from which ideal points may be extracted by data reduction. Bond and Messing (2015) apply SVD to their agreement matrix and extract the first left singular vector as a unidimensional representa-

tion of the ideology of each Facebook page. Due to the computational limitations of SVD applied to large rank matrices, I use an on-line variant of principal component analysis (PCA) in the data reduction stage (complete details provided in Appendix C.2). Because this data reduction method bears similarities to PCA, I depart from Bond and Messing (2015) by applying a final pre-processing step to the agreement matrix of subtracting off the mean from each column. This is a conventional step in pre-processing matrices for PCA as it prevents the first principal component from merely recovering the mean of the data.

C.2. *Online Dictionary Learning*

Though the original data reduction technique applied to networked social media data is singular value decomposition (SVD), the method is not scalable to large rank matrices. This problem is well-documented in computer science (Kobayashi et al. 2001; Mairal et al. 2010; Wu et al. 2017) where numerous methods have attempted to overcome the memory-constraints inherent in applying matrix factorization to large-rank matrices without sparsity.

Therefore instead of using SVD for data reduction, I employ the on-line dictionary learner (ODL) developed by Mairal et al. (2010). ODL approximates principal components of a matrix from i.i.d draws of individual vectors (or sets of vectors, η) from within that matrix, $\mathbf{G}^{p \times p}$. Though the reader is directed to Mairal et al. (2010) for technical details, I summarize the algorithm's main steps below.

Broadly, ODL follows Zou et al. (2006) in formulating principal component analysis (PCA) as a linear optimization problem and estimates the loadings in two stages: sparse coding and dictionary updating.

The sparse coding step occurs prior to the estimation of the principal components, wherein it (1) calculates the optimal effective sparsity of each principal component based on observed data and (2) represents the data as a sparse linear combination in order to reduce computational cost. Equation 2 describes the sparse coding step at iteration $t \in \{1, ..., T\}$, given a column vector j of data, $\mathbf{G}_{j,t}$, an initial matrix of z component loadings (the dictionary) $\mathbf{D}_{t-1}^{p \times k}$, and effective sparsity parameter, γ_{lasso} and γ_{ridge}

$$\boldsymbol{\theta}_t = \underset{\boldsymbol{\theta} \in \mathbb{R}}{\arg\min} ||\mathbf{G}_t - \mathbf{D}_{t-1}\boldsymbol{\theta}||_2 + \gamma_{lasso}||\boldsymbol{\theta}||_1 + \gamma_{ridge}||\boldsymbol{\theta}||_2. \tag{2}$$

Numerically, the least-squares problem presented by Equation 2 is evaluated via a Cholesky-based implementation of the LARS-Lasso algorithm (Efron et al. 2004, provides

deatiled explanations of the algorithm and its efficienct modifications). I choose not to induce sparsity when estimating loadings, hence the sparse coding equation simplifies to

$$\boldsymbol{\theta}_t = \underset{\boldsymbol{\theta} \in \mathbb{R}}{\arg\min} ||\mathbf{G}_t - \mathbf{D}_{t-1}\boldsymbol{\theta}||_2. \tag{3}$$

Upon calculating $\boldsymbol{\theta}_t$, the ODL algorithm stores all information necessary for the estimation of the principal component loadings in the updates of matrices $\mathbf{M}^{z \times z} = \mathbf{M}_{t-1} + \boldsymbol{\theta}_t \boldsymbol{\theta}_t'$ and $\mathbf{N}^{p \times z} = \mathbf{N}_{t-1} + \mathbf{G}_t \boldsymbol{\theta}_t'$.[4] This step permits the algorithm to discard all past information about $\mathbf{G}_t$ and $\boldsymbol{\theta}_t$ and simply maintain continual updates of $\mathbf{M}$, $\mathbf{N}$ and $\mathbf{D}$ in memory. These three matrices are sufficient for completing the update step of the ODL algorithm, wherein the approximations of the component loadings are updated according to Equation 3.

Next, the dictionary update step repeatedly iterates across each individual component loading, $k \in \{1, ..., K\}$, until convergence of the entire matrix of approximated loadings, $\mathbf{D}$. Formally, given the kth columns of matrices $\mathbf{M}$, $\mathbf{N}$, and $\mathbf{D}$, the unconstrained estimate of the component loadings is

$$\mathbf{c}_k = \frac{1}{\mathbf{M}_{k,k}} (\mathbf{N}_k - \mathbf{D}\mathbf{M}_k) + \mathbf{D}_k, \tag{4}$$

and the projection of the unconstrained loadings onto the constraint set is

$$\mathbf{D}_k = \underset{\mathbf{d} \in \mathbb{R}^k}{\arg\min} ||\mathbf{c}_k - \mathbf{d}||_2 \text{ such that } ||\mathbf{d}||_2 + \lambda_{lasso}||\mathbf{d}|| \leq 1. \tag{5}$$

Once $\mathbf{D}$ has converged, the ODL algorithm begins its next iteration by drawing a new i.i.d subsample of size η from the data and estimating a new $\boldsymbol{\theta}_t$. This process continues until completing the desired number of total iterations, T. Pseudo-code outlining the

[4] Both $\mathbf{M}$ and $\mathbf{N}$ are initiated as $\mathbf{0}$ matrices.

algorithm's steps as described above is available in Algorithm 1.

This iterative, observation-wise approach to estimating component loadings allows the application of PCA to large data sets which would otherwise be impossible to analyze because of an inability to fit into a computer's working memory. Since PCA, like SVD, offers a low-dimensional representation of political social media data, the ODL offers the opportunity to scale ideal point estimates for vast quantities of political communication.

I use the `SPAMS` (Mairal et al. 2017) `R` packages to implement online dictionary learning, running the algorithm for 5,000 re-sampling iterations (with 5 iterations at each dictionary update step) and by sampling 1,000 columns from the agreement matrix in each re-sampling. Once all iterations are complete, I divide each score by the standard deviation of all scores, center scores at the midpoint between their two modes, and subset the scores for posts authored in between November 11, 2015 and November 15, 2016. This completes the process of computing expressed ideology scores.

Algorithm 1: Online Dictionary Learning

Given:
Data $\mathbf{G}^{p \times p}$ with iid observations distributed $p(\mathbf{G})$
Regularization parameters γ_{lasso} and γ_{ridge}
Initial matrix of loadings $\mathbf{D}_0^{p \times K}$

1. Reset past information: $\mathbf{M}_0 \leftarrow \mathbf{0}^{K \times K}$, $\mathbf{N}_0 \leftarrow \mathbf{0}^{p \times K}$
2. For t in iterations 1 to T:
3. Subset $\mathbf{G}^{p \times \eta}$ by sampling η iid observations from $p(\mathbf{G})$
4. Optimize $\boldsymbol{\theta}_t^{K \times \eta}$ for fixed $\mathbf{D}_{t-1}$ (Equation 2)
5. $\mathbf{M_t} \leftarrow \mathbf{M}_{t-1} + \frac{1}{\eta} \sum_{n=1}^{\eta} \boldsymbol{\theta}_{t,n} \boldsymbol{\theta}'_{t,n}$, $\mathbf{N_t} \leftarrow \mathbf{N}_{t-1} + \frac{1}{\eta} \sum_{n=1}^{\eta} \mathbf{G}_{t,n} \boldsymbol{\theta}'_{t,n}$
6. Repeat
7. For k in dimensions 1 to K:
8. Evaluate $\mathbf{c}_k$ for fixed $\mathbf{M}$, $\mathbf{N}$, and $\mathbf{D}$ (Equation 4)
9. Optimize $\mathbf{D}_{t,k}$ for fixed $\mathbf{c}_k$ (equation 5)
10. Continue until convergence of $\mathbf{D}$
11. Return $\mathbf{D}_t$
12. Return $\mathbf{D}_T$

D. VALIDATION

This subsection of the measurement appendix provides additional evidence in favor of the validity of expressed ideology scores, as well as evidence that the network method of ideal point estimation is not biased by ancillary features of a post or its author. First, Appendix D.1 displays a random selection of candidates' Facebook posts and their corresponding expressed ideology scores to establish the face validity of the measure. Second, Appendix D.2 establishes content validity of expressed ideology scores with distributional evidence and a classification exercise that distinguishes between ideologically moderate and extreme posts. Third, I demonstrate evidence of the discriminant validity of the measure in detecting within-candidate difference in ideological expression (Appendix D.3). Fourth, Appendix D.4 presents further evidence of the measure's discriminant validity by analyzing within-state variation in expressed ideology scores. Fifth, I present evidence of the internal validity of the measure in Appendix D.5, which demonstrates that candidates' average scores remain consistent over time. Finally, I provide evidence of the stability of both users' ideology and posts' popularity to rule out non-rhetorical features of posts' social network as the cause of observed differences in expressed ideology scores among posts (Appendix D.6).

D.1. *Examples of Expressed Ideology*

Examination of scored Facebook posts demonstrates the face validity of the expressed ideology measure. Figure A2 displays nine scored posts and their corresponding expressed ideology scores. These nine posts were randomly selected from among 250 randomly selected posts that were used as a validation set. Additional details about the validation set are available in Appendix D.2. The posts shown in A2, are arranged with more liberally-scored posts towards the top and left, and with more conservatively-scored posts located towards the right and bottom.

Of the nine posts, Figure A2a receives the most extreme negative (liberal) score. The post, authored by Rep. Debbie Wasserman-Schultz (D-FL), expresses love and support for "the entire LGBT community" on the International Day Against Homophobia on Transphobia. The next-most negative post in the set (Figure A2b) was made by Rep. Barbara Lee (D-CA) and voices opposition to the displacement of Oakland's longtime residents as a contradiction to the "cultural diversity and artistic vibrancy that makes #Oakland" great. The most moderately scored post in the top row was produced by Rep. Bobby Scott (D-VA), which indicates that the Congressman enjoyed attending the Norfolk State University homecoming parade.

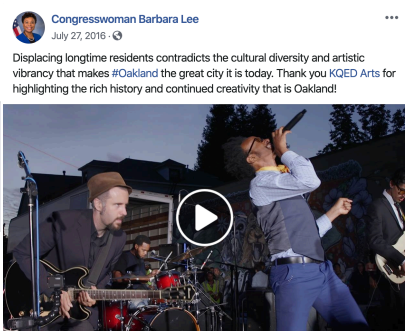
The second row of posts contains primarily moderate expressed ideology scores, which reflect the posts' largely bi-partisan or a-political content. In Figure A2d, Rep. Pete King (R-NY) described the suicide of an elderly veteran, and describes a bipartisan effort to investigate more effective mental health care for veterans. Rep. Robert Aderholt (R-AL) uploads a video of his experience attending a welcome home ceremony for combat troops (Figure A2e). And Rep. Sean Duffy (R-WI) posts a photo with his children dressed in Halloween costumes, one of whom is dressed as Donald Trump (Figure A2f).

More conservatively-scored posts are displayed in the bottom row of Figure A2. In the most moderate conservative post (Figure A2i), Rep. Diane Black (R-TN) links to a

YouTube video containing her Memorial Day well wishes as well as support for defense spending. A more positively-scored (conservative) post is display in Figure A2h. Rep. Dana Rohrabacher (R-CA) links to a Washington Times op-ed he authored, in which he calls for the prioritization of Christian immigrants from the Middle East. The most conservatively scored post in this sample (Figure A2g) is produced by Rep. Jeff Duncan (R-SC). The post outlines his bill, "The Terrorist Deportation Act," and link to a Miami Herald article describing the proposed bill at greater length.

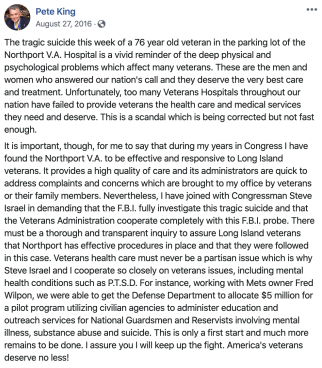
Though far from an exhaustive demonstration of posts, the correspondence between expressed ideology scores and the underlying content within each post is apparent. In terms of face validity, this exercise demonstrates that the scores are not only capable of distinguishing liberal, moderate, and conservative posts from one another, but are also able to reflect differences in the extremity within each ideological group.

(a) -1.05



(b) -0.54

(c) -0.12



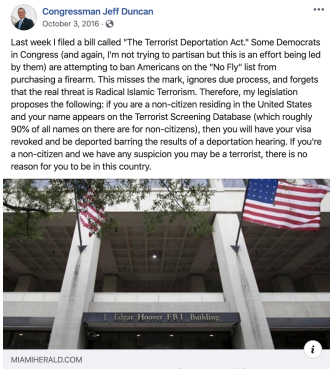
(d) 0.02

(e) 0.07

(f) 0.17

(g) 0.29

(h) 0.33



(i) 0.99

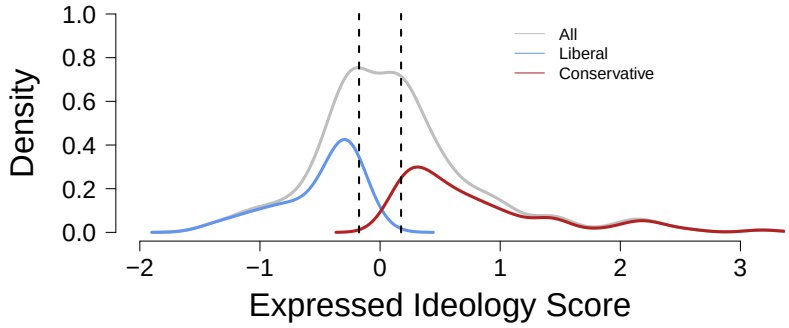
Figure A2: Sampling of Validated Facebook Posts

D.2. *The Distribution of Expressed Ideology*

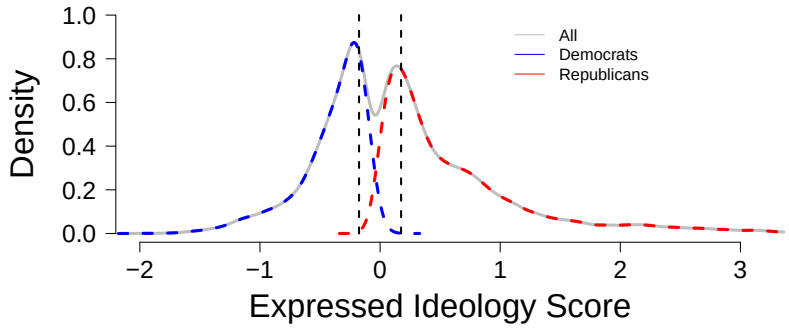
Though the previous subsection of the Validation Appendix provides qualitative evidence of the scores' validity, quantitative statistical analyses also support the validity of the expressed ideology scores. In this section I first provide distributional evidence that expressed ideology scores behave as one might expect given the partisan separation among legislators. Second, I performs a classification exercise using post scores and hand-labeled posts to attain a precise measure of the scores' accuracy.

To perform quantitative validation exercises, I randomly select a validation set of 250 scored Facebook posts. All posts in the validation set were used in the main text's analysis, and were hand-labeled by the researcher as liberal, moderate, or conservative. Posts are coded as moderate if they (a) contain a-political content, (b) express bi-partisan sentiments, (c) address low-politicized policy issues, (d) reference voter mobilization or get out the vote efforts. Liberal (conservative) posts are labelled based on (a) expressions of liberal (conservative) policy positions or (b) support for the Democratic (Republican) party or its members. Hand-labeling resulted in 72 liberal, 104 moderate, and 74 conservative posts.

Using the validation set, I present distributional evidence that expressed ideology scores are valid measures of candidates' communications. Figure A3a presents the density plot for the expressed ideology scores of all posts (grey), with density plots of posts coded as liberal (blue) and conservative (red) superimposed. The distributions of scores suggest that the measure does well in separating liberal and conservative posts from one another, as there are very few posts on the opposite side of 0 from their party's modal score. Additionally, the measure largely concentrates moderate posts near 0, while placing very few moderate posts in the ideologically-extreme tails.



(a) Validation Set



(b) Analysis Set

Figure A3: Distribution of Expressed Ideology Scores

Density plots for 250 randomly-selected, hand-labelled validation posts (a) and 27,430 posts used for empirical analysis (b). Colored lines in the top plot indicate the density for posts hand-labeled as liberal or conservative. Colored lines in the bottom plot indicate the density of posts authored by Democratic or Republican candidates. Dashed vertical lines are located at -0.175 and 0.175 to indicate the approximate threshold of moderate posts as used in the classification exercise. Posts with scores between these values are expected to be moderate, whereas those outside of the values are expected to be more extreme.

These trends are mirrored in the scores of posts more generally. Figure A3b plots the density of expressed ideology scores assigned to all posts used in the main text's analysis, with the densities of Democrats' post scores (blue) and Republicans' post scores (red)

superimposed. It is evident that the distribution of posts in the validation set is similar to that the larger corpus of posts. The similarity of the two distributions suggests that the validation set is representative of the broader set of posts, and the trends apparent in the validation set are likely to apply to the full set of analyzed posts as well.

To examine more precisely the scores' ability to distinguish between ideologically extreme versus moderate posts, I perform a classification exercise using the posts in the validation set described above. I collapse the expressed ideology scores assigned to these posts into a trichotomous variable that represents scores as Left, Right, or Center. These 3 bins are demarcated by cutoffs at -0.175 and 0.175, which are the approximate locations of the two modes located in the validation set's distribution of scores. These points are indicated on by vertical dashed lines in both Figures A3a and A3b. Posts with scores less than -0.175 are binned as Left, scores greater than 0.175 are binned as Right, and scores in between the cutoffs are binned as Center.

Table A5: Confusion Matrix of Expressed Ideology Scores as a Classifier

	Left (< -0.175)	Center ($< \|0.175\|$)	Right (> 0.175)
Liberal	0.931 (67)	0.069 (5)	0.000 (0)
Moderate	0.221 (23)	0.471 (49)	0.308 (32)
Conservative	0.000 (0)	0.081 (6)	0.919 (68)

Confusion matrix comparing hand-labeled ideology of 250 Facebook posts from the validation set against binned ideology scores for those posts. Cells contain the proportion of hand-labeled posts that fall into each bin based on expressed ideology scores. Raw counts expressed in parentheses.

The confusion matrix comparing the binned scores of these posts against their hand-labeled categories is presented in Table A5. For all three ideological categories, the binned expressed ideology scores correctly categorize the Facebook posts at rates far better than chance alone. For the 72 posts coded as liberal, 93% are correctly binned as Left based on their expressed ideology scores. Of the 5 liberal posts that are mis-classified, none are binned as Right. Similarly, of the 74 posts labeled as conservative, 92% are correctly

binned as Right based on their scores. Of the 6 conservative posts that are mis-classified, none are binned as Left. Finally, 47% of the 104 moderate-labeled posts were correctly binned in the Center category. Mis-classified moderate posts were binned as both Left and Right at roughly similar rates (22% and 31%, respectively). Broadly, the confusion matrix reiterates the trend described above, that expressed ideology scores are effective at distinguishing the ideological direction of a post's content, and are adequate in distinguishing the ideological extremity of posts. When the measure fails to assign an appropriate score to a post, it primarily does so by assigning a more extreme score to a moderate post.

D.3. *Within-Candidate Variation*

One concern with the network-based measure of post ideology is that the scores are detecting similarities and differences in candidates' relatively-static social media audiences, rather than mapping the variation in the expressed ideology of their rhetoric. If so, these scores should be no more dynamic than legislative-based measures of ideology, and will only change infrequently over time. To address this, I show that expressed ideology scores reliably detect ideological moderation within-candidates, precluding the possibility that differences in scores are solely attributable to static differences in audience.

I perform within-candidate validation using 20 random posts for each of four different candidates in the analysis. The candidates include two senators, Marco Rubio (R-FL) and Patty Murray (D-WA), and two House representatives, Paul Ryan (R-WI) and Nancy Pelosi (D-CA). For each candidate's set of 20 posts, I hand label the posts as moderate or partisan (using the same coding rules as those applied to the 150 validation posts, discussed in Appendix D.2) and compute the average expressed ideology score for each group. Results of this procedure are presented in Table A6. For all candidates, the posts coded as partisan are more ideologically extreme than those labeled as moderate. This implies that candidates' posts are not consistently receiving likes from the same set of followers. Instead it indicates variability in the pattern of likes that a candidate's posts receive, such that their ideologically moderate posts are, on average, being scored as more moderate than overtly partisan posts.

Table A6: Expressed Ideology Within Candidates

	Moderate	Partisan
P. Ryan	0.492 (8)	0.760 (12)
M. Rubio	0.619 (14)	0.707 (6)
P. Murray	-0.799 (7)	-0.809 (13)
N. Pelosi	-0.995 (6)	-1.172 (14)

Average expressed ideology scores for candidates' posts, labeled as moderate or partisan. Numbers in parentheses are total number of posts in the corresponding bin.

D.4. *Within-State Variation*

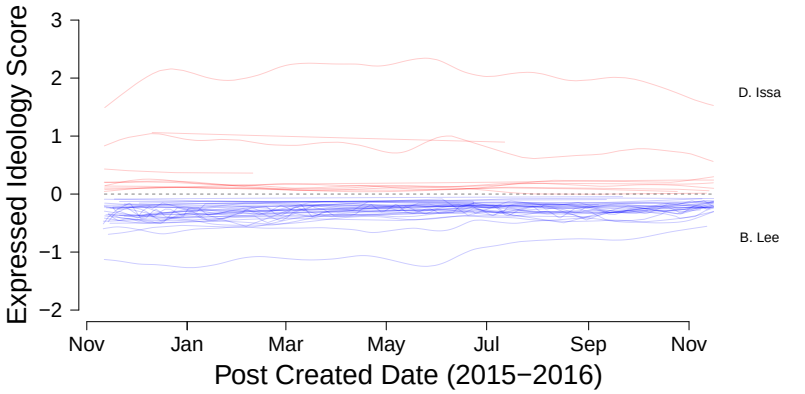
An additional concern surrounding a network-based measure of post-specific ideology is that the scores are responding to geographic regularities in user engagements. If social media users like the posts of their elected representatives, regardless of whether or not they support that representative's electoral bid, then differences in scores might indicate differences between states instead of differences in expressed ideology. Figure A4 visually depicts the variation of scores within three states — California, Florida, and Texas. These states are apt for comparing scores within and between states, because each state has numerous elected legislators and includes legislators of both parties. It is visually apparent from the plots that geography is not the primary determinant of a post's expressed ideology score. Legislators in all three states produces posts with scores across the same range, and are separated along party lines.

More rigorously, I estimate a two-way ANOVA on all posts included in the main analysis to determine how much of the variance in legislators' expressed ideology scores in explained by their state compared to their party. Table A7 presents the results of this ANOVA, which shows that while state and party both have a statistically significant effect on variation in expressed ideology scores ($p \approx 0.00$ for both groupings), the amount of variation explained by a legislator's party (mean square: 3,531.31) is much larger than that explained by the legislator's state (mean square: 137.71). These results lend support to the validity of expressed ideology scores, which, if they are reflective of candidates' ideological messaging, should be partly a function of a candidate's geographic characteristics, but should be substantially more affected by their partisan affiliation.

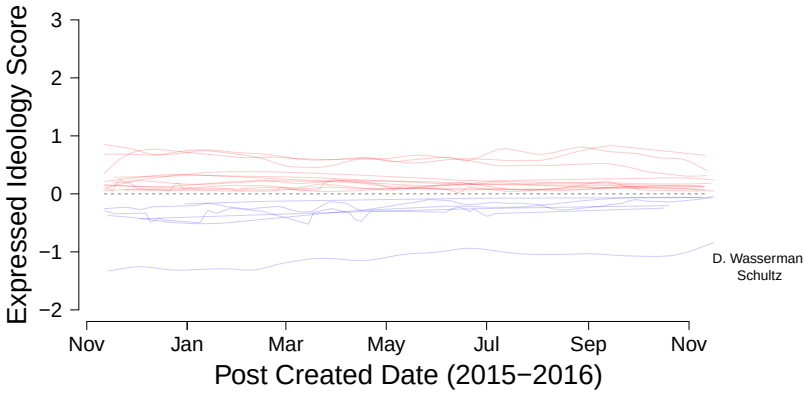
Table A7: ANOVA Comparing Variation Between States and Parties

	df	SS	MS	F	p-value)
State	45	6,197.11	137.71	684.22	0.00
Party	1	3,531.31	3,531.31	17,545.12	0.00
Residuals	27,383	5,511.38	0.20		

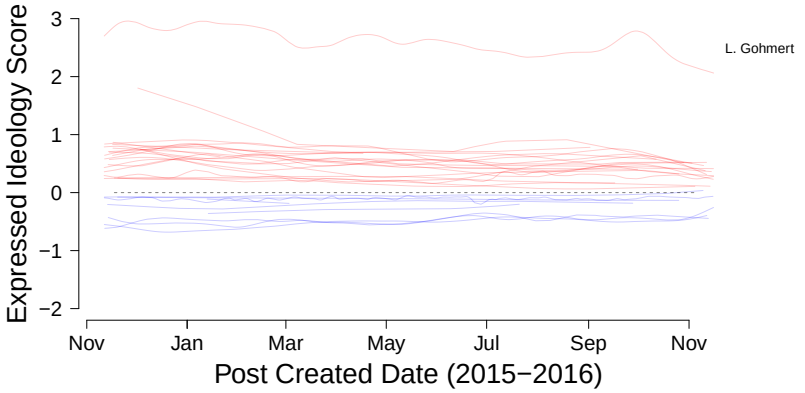
Results of an analysis of variance comparing variation in expressed ideology scores as a function of states and parties. Results show that both party and state are significant explanatory factors for expressed ideology scores, but that a legislator's partisanship explains far more of the variation in their expressed ideology scores than the state in which they are elected.



(a) California



(b) Florida



(c) Texas

Figure A4: Expressed Ideology Scores for Candidates in Four States

Trend lines (smoothed by cubic splines) of expressed ideology scores assigned to posts authored by Democratic (blue) and Republican (red) candidates in 3 states: California (a), Florida (b), and Texas (c). Trend line for Lamar Smith (R - TX) is excluded from the plots because he authors fewer than 4 posts in this period. Plots demonstrate considerable variation within candidates, within parties, and across states. In general, this variation suggests that expressed ideology scores are not an artifact of geographic regularities in the social network.

D.5. *Temporal Stability*

One concern with dynamic measures of ideology is the internal validity of the scores. To make causal claims from a time series of the measure requires demonstrating that the scores consistently measure similar content over time. To address this, I present the within-candidate correlations in average expressed ideology scores over 5 years (Table A8). The correlations between years are greater than 0.9 for all pairs of years in the five-year period. This high correlation between years speaks to the consistency of the measure.

Table A8: Temporal Correlation of Average Expressed Ideology Scores

	2014	2015	2016	2017
2013	0.971	0.954	0.936	0.908
2014		0.979	0.961	0.935
2015			0.983	0.958
2016				0.975

Pearson correlation of candidates' yearly average expressed ideology scores. Sample includes 281 candidates with scored Facebook posts in the years from 2013 to 2017.

D.6. *Changing Rhetoric or a Changing Network?*

An inherent ambiguity in this network-based measure of expressed ideology is the source of variation in the composition of likes that each post receives. Ideally, the composition of users who endorse any given post is determined by the expressed ideology in the post as well as the corresponding reaction of users to that ideological content. However, there may exits other mediators that determine the composition of a post's likes by manipulating the post's exposure across the Facebook platform. The relative exposure of certain posts could be affected by platform-wide algorithmic changes, temporal variation in the number of active Facebook users, or candidates who pay to promote posts in users' feeds. In such cases, the differences in likes across candidates' posts would not be due users' responses to candidate rhetoric, but would instead reflect differences in the baseline supply of likes attributed to the affected posts.

The main safeguard against non-ideological network regularities biasing inference is the difference-in-differences-like design used to analyze this data. The linear two-way fixed effects (TWFE) model is difference-in-differences-like in the sense that it controls for unobserved confounders that affect candidates' outcomes beyond treatment status alone. On the one hand, the time fixed-effects control for temporal confounding (such as Facebook's number of active users or national political events) that might diferentially affect engagement activity across the Facebook platform in a given month. On the other hand, because I compare post-primary changes in expressed ideology within the *primary* group against the baseline post-primary ideological moderation exhibited by the *no primary* group, this also means that platform-wide changes across time periods affecting all candidates should be controlled. Thus, any time-dependent or platform-wide changes in the relative popularity of posts are controlled for by the TWFE design through its simultaneous use of time fixed-effects and cross-group comparisons of post-treatment outcomes.

However, though the TWFE safeguards against biased inference, it does not prove

that the expressed ideology measure is immune to fluctuations in user activity that may be independent of candidates' rhetoric. To directly address whether network features affect measurement of the dependent variable, I present evidence that users' endorsement behavior remains relatively stable over the period of study, while the ideological extremity of users who endorse the posts used in the analysis decreases over time. This pattern of stable user endorsements and moderating user ideology is more consistent with the ideal, reaction-based model of user endorsement decisions, relative to alternate explanations for the network's structure. To first establish the stability of users' endorsement activity, I present the proportion of unique users who like posts during each week in the year preceding the 2016 general election. Specifically, Figure A5 shows the proportion of unique users (among all users who like posts used in the analysis) that endorse posts in each week, separated by *primary* (black) and *no primary* (grey) groups. On average, approximately 3% of all unique users in the sample like at least one candidate's post per week. Apart from a spike in the number of engagements in June and November of 2016, the amount of weekly user endorsement activity remains relatively stable.[5] This stability in the number of unique users that engage with candidates' Facebook posts suggests that the observed changes in expressed ideology scores are not due to a change in the overall user activity on the platform. This rules out secular trends in user activity as a potential source of the observed moderation in expressed ideology scores over time. Importantly, even during the June and July activity spike, both the *primary* and *no primary* groups show an increase in user engagements during the spike, suggesting that the TWFE model will appropriately control for such period-specific shocks.

[5]The June 2016 spike in unique users coincides with large numbers of engagements on posts addressing the June 12 Pulse Nightclub mass shooting in Orlando, FL and the subsequent sit-in by House Democrats on June 22 to demand gun reform legislation. November 2016's spike coincides with engagements on posts in the aftermath of the 2016 election.

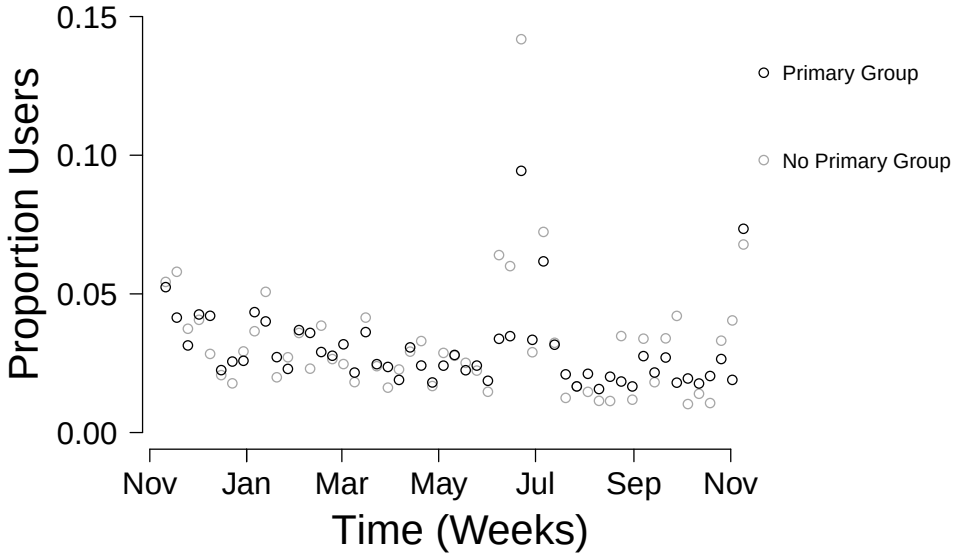


Figure A5: Engagements from Unique Users over Time

The proportion of unique users in the sample who like a candidate's post in each week from November 11, 2015 to November 15, 2016. For the both the *primary* (black) and *no primary* (grey) groups, the proportion of total users who engage with politicians' content remains relatively stable (approximately 3%) throughout the period of study except for spikes in June and November of 2016.

Second, I show that despite the stability of users' aggregate engagement activity, ideologically extreme users endorse fewer posts later in the election cycle. I calculate a user's ideology as the average expressed ideology score for all posts which they like during the period of analysis. Bond and Messing (2015) use a similar method for scaling the users in their analysis, instead using the average page-level ideology score for all of the political pages that a user likes. They validate this approach by comparing these scores against users' self-reported ideology and find a discernible positive relationship between the two measures. As with the post-level scores, I evaluate the absolute value of users' average ideology scores in order to measure the overall extremity of their endorsement activity. Figure A6 shows the number of unique users who engage with candidates' content during the electoral period. Extreme users (grey) are those with ideological extremity scores at or above the 90th percentile of user scores, and moderate users (black)

are those with scores beneath the 90th percentile. The trend lines are smoothed with cubic splines. The trend line for moderate users remains relatively stable over time, which is consistent with the overall amount of engagement activity throughout the period as depicted in Figure A5. However the trend for extreme users decreases throughout the electoral period, indicating that extreme users are not applying as many likes to content that candidates produce nearer to the general election. This evidence is consistent with the expectations set forth in describing the measurement method. The appearance of more moderate Facebook posts over time during the 2016 election corresponds with decreasing engagement from extreme users who endorsed content earlier in the electoral cycle. This pattern is consistent with occurrence of actual moderation among candidates posts over time, rather than artifacts within the network causing the moderating trend in expressed ideology scores.

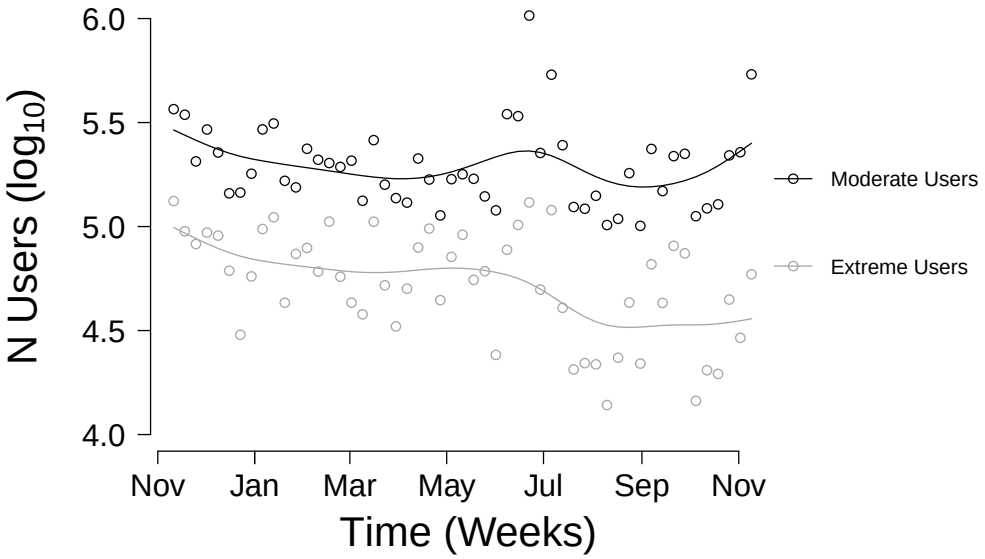


Figure A6: User Extremity over Time

The number of unique users in the sample, separated by ideological extremism, who like a candidate's post in each week from November 11, 2015 to November 15, 2016. Trend lines estimated with a cubic spline. Moderate users (black), those with ideological extremism scores below the 90th percentile of scores, engage with candidates' content at relatively stable rates throughout the election. Ideologically extreme users (grey), with ideological extremism at or above the 90th percentile, decreasingly endorse content produced by candidates as the general election draws closer.

E. IMPUTATION

This section of the appendix details the imputation method used for missing observations in the data set used in the text's main analysis. Though the average candidate authors more than one Facebook posts per week (Appendix B, Table A1), the weekly time series of candidates' expressed ideological extremity contains substantial missingness. Figure A7 shows missingness among units over time for candidates in the *primary* and *no primary* groups. Nearly half of all candidate-week observations are missing.

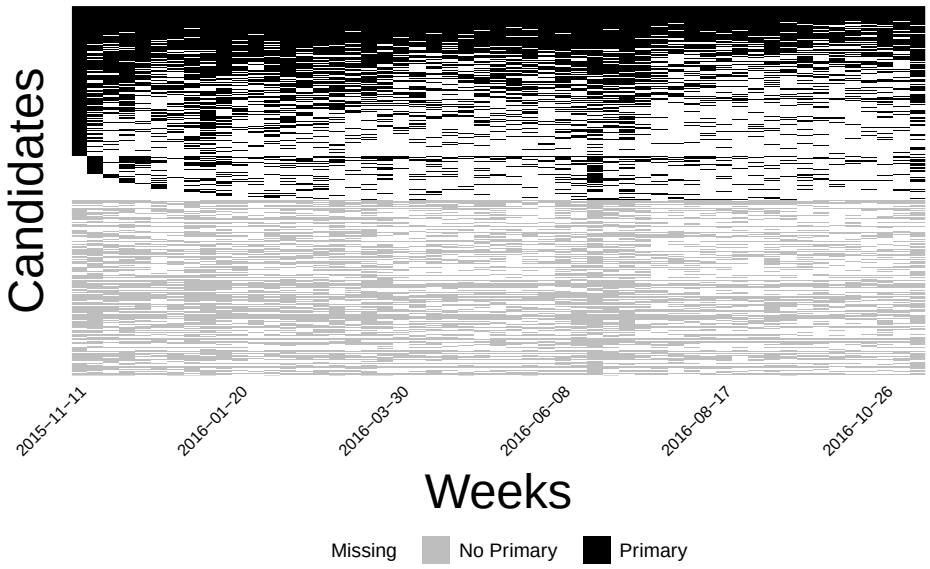


Figure A7: Missingness in Expressed Ideological Extremity Over Time

Missing and observed values in expressed ideological extremity for candidate-weeks in the main analysis. Each row represents a candidate, whose treatment status is indicated by the shading of the cell. The first week begins on November 11, 2015 and the final week ends on November 15, 2016. Plot produced with the `PanelView` (Liu and Xu 2019) `R` package.

A large portion of missing observations occur because the expressed ideology measure censors posts which receive fewer than 100 likes. However, this does not mean that missingness is non-ignorable. So long as the process that determines whether posts receive less than 100 likes is not due to systematic unobservables, censored observations can be considered missing at random (MAR), and can be imputed from observed data.

To rectify the missingness in the data I use multiple-imputation for time series as specified in Honaker and King (2010) with the aid of the `AmeliaII` (Honaker et al. 2011) `R` package. To impute expressed ideological extremity I use three candidate-level characteristics: (a) candidates' primary cohort, (b) candidates' treatment grouping, and (c) candidates' chamber of Congress to which they are subsequently elected. In addition to these characteristics, I apply candidate-specific linear trend to account for candidate- and time-specific effects on expressed ideological extremity. I impute 10 datasets and estimate each empirical model once for each imputed dataset, averaging the resulting coefficients, and computing the standard error estimates as directed in Honaker and King (2010). This process of censoring and imputing is preferable to estimating expressed ideology scores for posts with few endorsements, as posts with less than 100 endorsements can produce unstable estimates that have little relationship with the underlying post content.

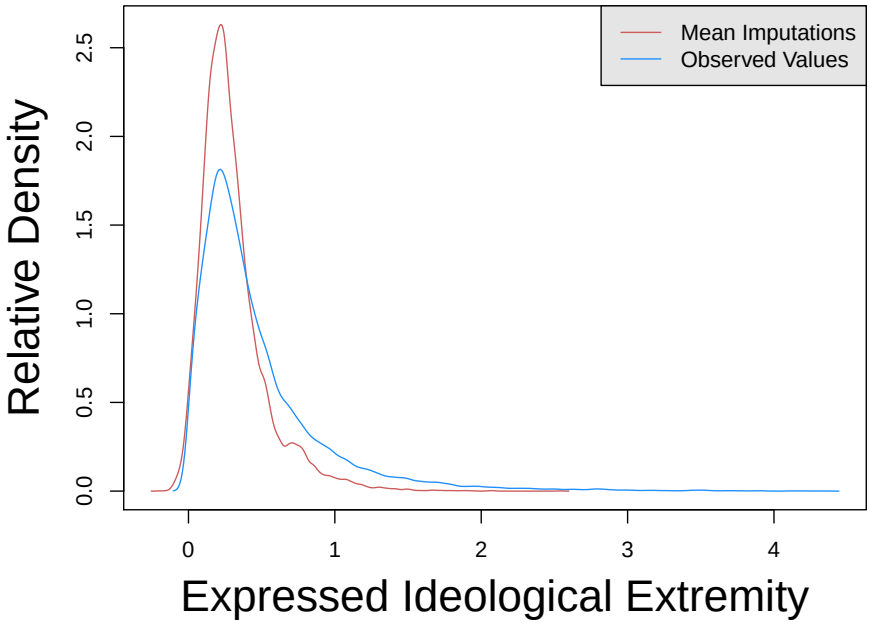


Figure A8: Imputation of Expressed Ideological Extremity

Density of observed (blue) and imputed (red) expressed ideological extremity. The approximate overlap of the two densities suggests that imputed scores are not substantially different from those observed in the data. Plot produced with the `AmeliaII` (Honaker et al. 2011) `R` package.

Figure A8 plots the densities of observed (blue) expressed ideological extremity scores against those imputed across the 10 imputed datasets (red). The close overlap between the two densities suggests that the multiple imputation model successfully approximates scores that resemble those observed in the data.

F. ROBUSTNESS

This section of the appendix verifies that the results presented in the main text are robust to alternate modeling decisions. As robustness checks, I first present the complete set of expressed ideology trends across primary cohorts (Appendix F.1) to demonstrate that the trends depicted in the main text (Figure 2) are consistent with the majority of cohorts in the analysis. Second, I perform a balance test to rule out potential differences across *primary* and *no primary* groups apart from treatment assignment alone. Next, I conduct a unit root analysis in Appendix F.3 to demonstrate that the TWFE results are not erroneously detecting non-stationary time trends in the data, but rather represent actual changes in observed outcomes due to primary elections. Finally, Appendix F.5 presents statistical results using estimators other than the linear two-way fixed effects model to test how results hold up under alternate assumptions.

F.1. *Trends in Expressed Ideological Extremity by Primary Cohort*

The main text presents visual evidence of post-primary moderation's effect on select primary cohorts (Figure 2) to illustrate the finding reported in the two-way fixed effects model. This section of the appendix presents the trend lines for all primary cohorts in Figure A9 so that interested readers might assess the effect of treatment for the complete set of candidates. In all plots, solid black lines indicate observed average weekly expressed ideological extremity scores and dashed black lines indicate the counterfactual trend that one would expect from the cohort had it not faced primary election (inferred by the *no primary* group's trend line depicted in grey). Across all 16 cohorts, 12 demonstrate visual evidence of post-primary moderation consistent with the findings of the TWFE model.

Another notable feature of the trend lines is that for all but three cohorts, the trends in pre-treatment outcomes are approximately parallel to the trend line of the *no primary* group. This supports the notion, asserted in the main text, that TWFE approximates a multi-period difference-in-differences (DID) estimator (Goodman-Bacon 2018). A critical assumption in DID analyses is that the average trends in outcomes among units in both treatment and control groups would be parallel in the absence of treatment intervention. While there is no precise method for confirming whether parallel trends exist, visual inspection of the outcomes for *primary* and *no primary* groups in Figure A9 suggests that this assumption is at least plausible for most primary cohorts in the analysis.

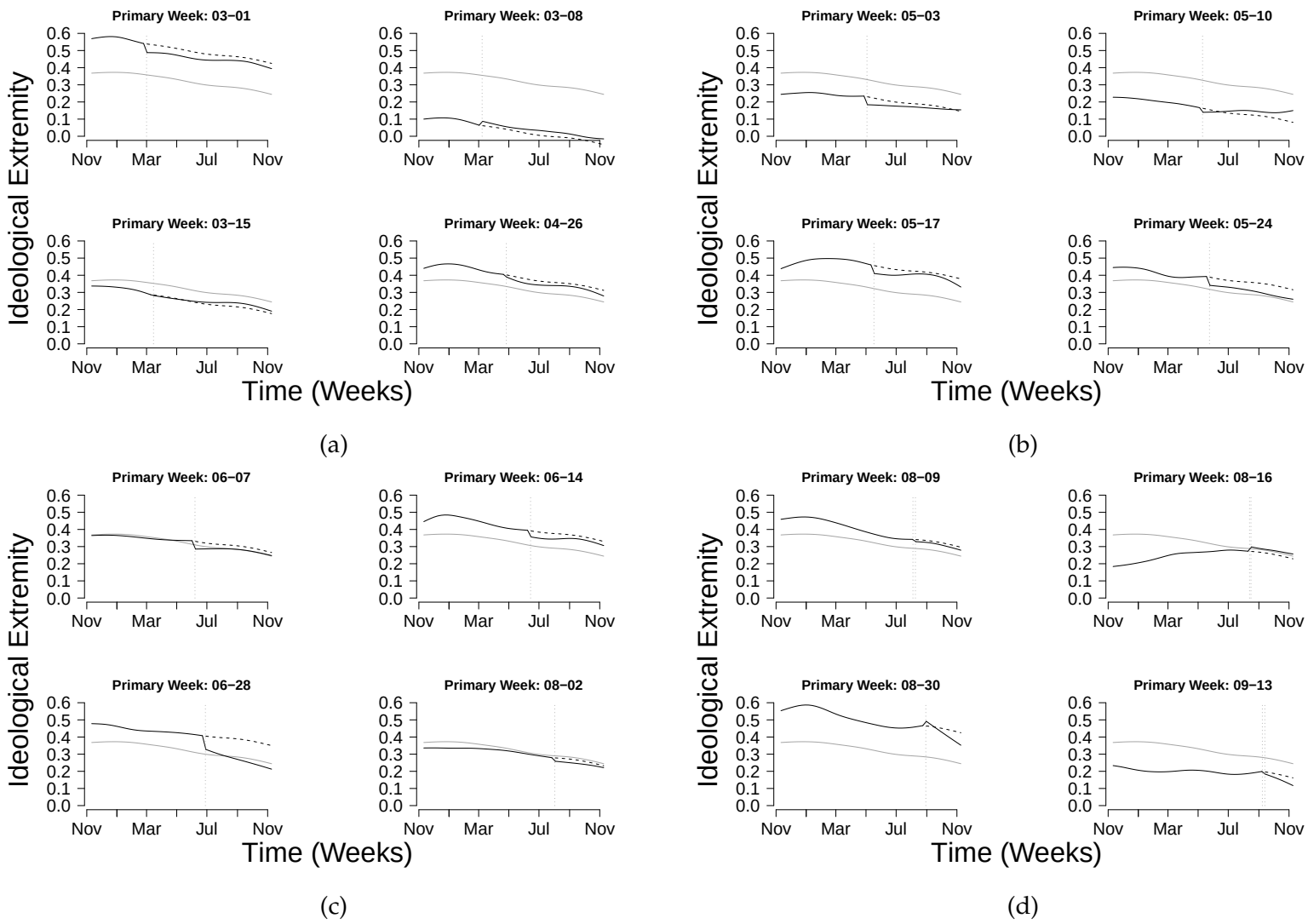


Figure A9: Trends in Expressed Ideological Extremity by Primary Cohort

Average expressed ideological extremity by primary cohort, smoothed by cubic spline. Grey line represents the average expressed ideological extremity of the *no primary* group. Solid (dashed) black lines indicate observed (counterfactual) trends in ideological extremity for the *primary* group. Underlying data consists of average values in expressed ideological extremity from 10 multiply-imputed datasets. Time series ranges from November 11, 2015 to November 15, 2016.

F.2. *Balance Test for Primary and No Primary Groups*

Conducting a difference-in-differences-like analysis with linear two-way fixed effects (TWFE) is advantageous because it permits a causal interpretation of the treatment effect of primary elections on expressed ideological moderation. One of the requirements for this causal interpretation to hold is that the characteristics of the treatment (*primary*) and control (*no primary*) groups should differ only with respect to treatment assignment. However, given that the absence or presence of primary competition may results from strategic actions taken by candidates in the pre-primary period, it could be that the assumed exogeneity of treatment assignment is implausible. To examine whether there exist differences between the two groups that could indicate endogenous causes of primary competition, I run a logistic regression with treatment group as the outcome and include candidates' party, chamber, incumbency status, and ideological extremity (measured as the absolute value of their DW-NOMINATE score) as predictors. The results, presented in Figure A9, fail to find any statistically distinguishable differences between groups. This suggests that the assumed similarity of the groups may not be implausible. In particular, the fact that ideological extremity does not significantly differ across groups indicates that candidates in this sample are not able to select into *primary* and *no primary* groups based on ideology alone.

Table A9: Balance Test for Primary Competition

Ideo. Extremity	1.05
	(0.80)
Republican	0.01
	(0.22)
Senate	0.10
	(0.40)
Incumbent	-1.29
	(1.16)
(Intercept)	0.90
	(1.20)
N	375

Coefficients with bold text are statistically significant at the 0.05 level. Standard errors in parentheses.

F.3. *Testing for Non-Stationarity and Unit Root*

One potential concern surrounding linear two-way fixed effects (TWFE) models is that non-stationarity in the time series might bias the estimated treatment effect. Because the trend lines for expressed ideology (presented in full in Appendix F.1) exhibit a secular downward trend for all groups of candidates, this raises cause for concern. To formally determine whether the time series is non-stationary, I employ the unit-root test recommended by Phillips and Moon (2000) and utilized by Caughey et al. (2017, pg. A-14). The test involves regressing the analytical model's covariates and lagged dependent variable on the first-differenced outcome variables. Formally the equation is:

$$\Delta y_{it} = \beta D_{it} + (\rho_l - 1)\, y_{i,t-1} + \alpha_i + \omega_t + \epsilon_{it} \tag{6}$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$ and the other notation is consistent with that presented in Equation 1 in the main text. If the coefficient on the lagged dependent variable $(\rho_l - 1)$ is statistically distinguishable from 0, it indicates that ρ_1 is not equal to 1, and hence the time series does not have a unit root. The regression results are presented in Table A10. The coefficient on the lagged dependent variable is negative and statistically significant (-0.87, p: 0.02), indicating that a unit root is unlikely. Moreover, the effect of treatment is comparable to that reported in the main analysis (Table 1), suggesting that regardless of whether the outcome is analyzed raw or differenced, the effect of treatment is consistent.

Table A10: Unit Root Test

Treatment	**-0.02**
	(0.01)
Ideo. Extremity $t-1$	**-0.87**
	(0.02)
Candidate & Week FEs	✓
N	19,875

Coefficients with bold text are statistically significant at the 0.05 level. Standard errors in parentheses.

F.4. *Alternate Lag Specifications*

The analysis presented in the main text uses a linear two-way fixed effects (TWFE) model with a one-period lagged dependent variable. As in Caughey et al. (2017), I lag the dependent variable to explicitly model the effect of past outcomes on treatment status. In this appendix section, I present the results of linear TWFE models that include alternate numbers of lagged dependent variables to verify that the results are not driven by the choice of a one-period lag. Each regression is of the form reported in Equation 1 of the main text and uses panel-corrected standard errors (Bailey and Katz 2011) to account for spherical errors over time and across candidates. Table A11 reports the results of each using a series of 0 (column 1) to 3 (column 4) lagged dependent variables in the regression. It is evident that the results are robust to varying lag specifications, with the statistical significance, sign, and relative magnitude of treatment remaining consistent across all columns of the table.

Table A11: Two-way Fixed Effects Results

Lags	(0)	(1)	(2)	(3)
Treatment	**-0.023**	**-0.021**	**-0.019**	**-0.018**
	(0.005)	(0.005)	(0.005)	(0.005)
Ideo. Extremity $t-1$		**0.126**	**0.114**	**0.107**
		(0.025)	(0.023)	(0.023)
Ideo. Extremity $t-2$			**0.096**	**0.088**
			(0.023)	(0.022)
Ideo. Extremity $t-3$				**0.078**
				(0.022)
Candidate & Week FEs	✓	✓	✓	✓
N Candidates	375	375	375	375
N Observations	19,875	19,875	19,875	19,875

Linear two-way fixed effects regression results. Column headers indicate the number of lagged dependent variables included in the model. Panel-corrected standard errors in parentheses. Coefficients with bold text are statistically significant at the 0.05 level.

F.5. *Alternate Estimators*

To confirm that the results in the main analysis are not an artifact of the modeling assumptions in linear two-way fixed effects regression (TWFE), I conduct alternate analyses that relax some of these assumptions. First, I use a first-difference (FD) analysis using the non-parametric fixed effects estimator proposed in Imai and Kim (2019). The estimator controls for time-invariant confounding by comparing matched treated and controlled observations within-units over time. Though this approach does not explicitly model time-specific effects on the outcome, its non-parametric nature relaxes the linearity assumption in TWFE that might erroneously drive results when linearity is not a property of the true data generating process.

Second, I use the generalized synthetic control (GSC) method proposed by Xu (2017). GSC performs a multi-period difference-in-differences (DID) estimation where, first, each unit's past outcomes are used to predict post-treatment outcomes, and predicted outcomes are compared to actual post-treatment outcomes. The benefit of this counterfactual estimation technique is that it relaxes the parallel trends assumption required for identification in standard DID models as it allows each unit to predict its own counterfactual post-treatment outcomes.

For each alternate estimator, I directly estimate the effect of treatment on expressed ideological extremity. No additional controls or lags are included. As in the main analysis, each model is applied to the 10 imputed datasets, and results are averaged according to the method of Honaker and King (2010). Across each imputed dataset, the GSC model uses a one-factor model — deemed optimal by leave-one-out cross-validation — to predict counterfactuals. I fit the relevant models using `wfe` (Kim and Imai 2019) and `gsynth` (Xu and Liu 2018) `R` packages, respectively. The result of these models are presented in Table A12 and are largely consistent with those presented in the main analysis. For the non-parametric FD model, the sign and statistical significance of treatment (coef: -0.04,

p: 0.02) is consistent with that of linear TWFE. For each of the models applied to the 10 imputed datasets, I perform a White's non-specification test (White 1980), which operates under the null hypothesis that a linear fixed effects model correctly models the data. Without a principled approach to aggregating p-values across models, I instead report the proportion of non-specification tests that reject the null hypothesis at the 0.05 level. All 10 non-specification tests reject the null hypothesis, suggesting that the non-parametric FD design may be more appropriate for modeling within-unit treatment effects when relaxing the linearity assumption and excluding the time-specific controls of linear TWFE.

Table A12: Alternate Model Specifications

	(FD)	(GSC)
Treatment	**-0.04**	**-0.03**
	(0.02)	(0.02)
Candidate FEs	✓	✓
Week FEs		✓
N Candidates	375	375
N Observations	19,223	19,223
Proportion White's p Rejected	1.00	

Alternate time series causal inference estimators. All results are averaged across 10 multiply-imputed datasets. Coefficients with bold text are statistically significant at the 0.05 level. Standard errors in parentheses. Non-parametric model (column 1) uses heteroscedastic / autocorrelation robust standard errors. Generalized synthetic control model (column 2) uses non-parametric bootstrapped standard errors from 1,000 bootstrapped samples.

Similarly, the GSC results also support the findings in the main analysis. The ATT estimate is negative, statistically significant, and similar in magnitude (coef: -0.3, p: 0.04) to that of linear TWFE. This finding lends support to the idea that comparing candidates in competitive primaries against those without primary competition offers a DID-like estimate of treatment. Even though the GSC relaxes the parallel trends assumption, the model results are similar to the linear TWFE results presented in the main text, which would otherwise require the presence of linear parallel trends for causal identification.

As a final note, another candidate model for this analysis is the non-parametric two-

way fixed effects model which extends from the FD model described above (Imai and Kim ND). The estimator controls for both unit-specific and time-specific confounding and is equivalent to a multi-period difference-in-differences (DID) estimator whereby matched sets are constructed for both treated and controlled units in each period and the differences between groups are evaluated. Though the model only requires the parallel trends assumption for causal inference, relaxing the linearity assumption in the is context is inappropriate. As with the non-parametric first-difference model, I compute a White's p statistic for each of the 10 imputed data sets using the non-parametric TWFE model. 9 of the 10 tests fail to reject the null hypothesis, suggesting that non-parametric TWFE is no better in modeling the data than the linear TWFE approach. As such, I do not model the data using non-parametric TWFE.

G. APPENDIX REFERENCES

References

Bailey, Delia and Jonathan N Katz (2011). "Implementing panel corrected standard errors in R: the pcse package". *Journal of Statistical Software* 42.CS1, pp. 1–11.

Bond, Robert and Solomon Messing (2015). ""Quantifying social media's political space: Estimating ideology from publicly revealed preferences on Facebook"". *American Political Science Review* 109.1, pp. 62–78.

Caughey, Devin, Yiqing Xu, and Christopher Warshaw (2017). "Incremental democracy: The policy effects of partisan control of state government". *The Journal of Politics* 79.4, pp. 1342–1358.

Efron, Bradley et al. (2004). "Least angle regression". *The Annals of statistics* 32.2, pp. 407–499.

Goodman-Bacon, Andrew (2018). *Difference-in-Differences with Variation in Treatment Timing*.

Honaker, James and Gary King (2010). "What to do about missing values in time-series cross-section data". *American journal of political science* 54.2, pp. 561–581.

Honaker, James, Gary King, and Matthew Blackwell (2011). "Amelia II: A Program for Missing Data". *Journal of Statistical Software* 45.7, pp. 1–47. URL: `http://www.jstatsoft.org/v45/i07/`.

Imai, Kosuke and In Song Kim (2019). "When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data?" *American Journal of Political Science* 63.2, pp. 467–490.

— (ND). "On the Use of Two-way Fixed Effects Regression Models for Causal Inference with Panel Data". *Working Paper*.

Kim, In Song and Kosuke Imai (2019). *wfe: Weighted Linear Fixed Effects Regression Models for Causal Inference*. R package version 1.9.1. URL: `https://CRAN.R-project.org/package=wfe`.

Kobayashi, Mei et al. (2001). "Estimation of singular values of very large matrices using random sampling". *Computers & Mathematics with Applications* 42.10-11, pp. 1331–1352.

Liu, Licheng and Yiqing Xu (2019). *panelView: Visualizing Panel Data*. R package version 1.1.2. URL: `https://CRAN.R-project.org/package=panelView`.

Mairal, Julien et al. (2010). "Online learning for matrix factorization and sparse coding". *Journal of Machine Learning Research* 11.Jan, pp. 19–60.

Mairal, Julien et al. (2017). *SPAMS: a SPArse Modeling Software*. R package version 2.6.2. URL: `http://spams-devel.gforge.inria.fr/`.

Phillips, Peter CB and Hyungsik R Moon (2000). "Nonstationary panel data analysis: an overview of some recent developments". *Econometric reviews* 19.3, pp. 263–286.

White, Halbert (1980). "Using least squares to approximate unknown regression functions". *International Economic Review*, pp. 149–170.

Wu, Lingfei, Eloy Romero, and Andreas Stathopoulos (2017). "Primme_svds: A high-performance preconditioned svd solver for accurate large-scale computations". *SIAM Journal on Scientific Computing* 39.5, S248–S271.

Xu, Yiqing (2017). "Generalized synthetic control method: Causal inference with interactive fixed effects models". *Political Analysis* 25.1, pp. 57–76.

Xu, Yiqing and Licheng Liu (2018). *gsynth: Generalized Synthetic Control Method*. R package version 1.0.9. URL: `https://CRAN.R-project.org/package=gsynth`.

Zou, Hui, Trevor Hastie, and Robert Tibshirani (2006). "Sparse principal component analysis". *Journal of computational and graphical statistics* 15.2, pp. 265–286.