Supporting Information (SI)

Greater Internet use is not associated with faster growth in political polarization among US demographic groups

Levi Boxell, *Stanford University**
Matthew Gentzkow, *Stanford University and NBER*
Jesse M. Shapiro, *Brown University and NBER*

December 2017

List of Tables

List of Figures

*E-mail: lboxell@stanford.edu, gentzkow@stanford.edu, jesse_shapiro_1@brown.edu

Contents

SI Table S1: Correlations among measures of polarization

	Partisan affect	Ideological affect	Partisan sorting	Partisan-ideology	Perceived partisan-ideology	Issue consistency	Issue divergence	Straight-ticket	Index
Partisan affect	1.000								
Ideological affect	0.582	1.000							
Partisan sorting	0.927	0.661	1.000						
Partisan-ideology	0.926	0.694	0.980	1.000					
Perceived partisan-ideology	0.786	0.706	0.735	0.819	1.000				
Issue consistency	0.858	0.936	0.983	0.952	0.783	1.000			
Issue divergence	0.746	0.947	0.920	0.891	0.774	0.946	1.000		
Straight-ticket	0.882	0.687	0.938	0.964	0.842	0.918	0.931	1.000	
Index	0.876	0.818	0.949	0.975	0.872	0.974	0.950	0.960	1.000

Notes: Table shows the correlation across the eight polarization measures and the index. The correlation is calculated using pairwise-complete observations of each polarization measure m_t across presidential election years from 1972 to 2016.

SI Table S2: Predicted internet, 1996 – 2016

	Estimator: Weighted least squares Dependent variable: Internet use
Year: 1996	0.400
	(0.017)
Year: 2000	0.753
	(0.017)
Year: 2004	0.839
	(0.018)
Year: 2008	0.884
	(0.017)
Year: 2012	1.014
	(0.017)
Year: 2016	1.023
	(0.018)
Age Group: 40-64	-0.070
	(0.008)
Age Group: 65+	-0.311
	(0.011)
Gender: Male	0.006
	(0.007)
Race: Hispanic	0.073
	(0.016)
Race: Other	0.112
	(0.020)
Race: White	0.125
	(0.012)
Education: Grade School	-0.510
	(0.023)
Education: High School	-0.272
	(0.009)
Education: Some College	-0.075
	(0.010)
Region: South	0.002
	(0.008)
N	9351
R^2	0.816

Notes: Table shows the coefficients from a weighted least squares regression. Weights are the ANES survey weights. For estimation, the sample is restricted to respondents who have valid responses to the questions needed to construct each independent and dependent variable. For prediction, the sample is restricted to respondents who have valid responses to the questions needed to construct each independent variable. Dependent variable is an indicator for whether an individual uses the internet taken from the ANES (see section 1 of the SI appendix for details on the variable construction). All covariates are indicator variables. Age groups are defined using *VCF0101* (for 1996–2012) and *V161267* (for 2016), gender is defined using *VCF0104* (for 1996–2012) and *V161267* (for 2016), race is defined using *VCF0105b* (for 1996–2012) and *V161310x* (for 2016), education is defined using *VCF0110* (for 1996–2012) and *V161270* (for 2016), and region is defined using *VCF0113* (for 1996–2012) and *V161010d* (for 2016). We treat the response '95. Other SPECIFY' as an invalid response to *V161270*. Standard errors are in parentheses.

SI Table S3: Linear model coefficients

Model:	(1)	(2)	(3)	(4)	(5)
	No internet effect	Internet use (ANES)	Internet use (Pew)	Campaign information (ANES)	Social media (Pew)
Age Group: 18-39	0.882	0.869	0.976	0.933	0.884
	(0.044)	(0.120)	(0.118)	(0.086)	(0.043)
Age Group: 40-64	0.987	0.975	1.052	1.018	0.967
	(0.044)	(0.106)	(0.088)	(0.064)	(0.046)
Age Group: 65+	0.974	0.970	0.981	0.964	0.935
	(0.044)	(0.054)	(0.045)	(0.047)	(0.053)
Time	0.016	0.015	0.023	0.021	0.022
	(0.003)	(0.008)	(0.008)	(0.008)	(0.005)
Internet	—	0.029	-0.204	-0.204	-0.152
		(0.237)	(0.236)	(0.297)	(0.125)
N	18	18	18	18	18
R^2	0.996	0.996	0.996	0.996	0.996

Notes: Table shows the coefficients from an OLS regression of the polarization index on a time trend (measured in years relative to 1996), age group indicators, and a measure of internet or social media use. The unit of analysis is the age group and year. The sample period consists of presidential election years from 1996 to 2016, and the age groups are the 18–39, 40–64, and 65+ age groups. Five different models are estimated. The first model excludes the internet use variable. The second model uses the ANES internet use variable. The third model uses the the Pew Research Center internet use variable. The fourth model uses the ANES obtaining campaign information online variable. The fifth model uses the Pew Research Center social media use variable. Standard errors are in parentheses.

SI Table S4: Sensitivity analysis for proportion of linear trend explained by internet

| Model	$\left(\hat{\beta}_c - \hat{\beta}\right)/\hat{\beta}_c$	95% CI
Heterogenous impact		
Internet use: ANES	0.255	(-0.292, 0.802)
Internet use: Pew	-0.168	(-0.628, 0.292)
Campaign information: ANES	-0.151	(-0.693, 0.39)
Social media: Pew	-0.303	(-0.631, 0.025)
Social spillovers		
Internet use: ANES	0.078	(-0.458, 0.614)
Internet use: Pew	-0.327	(-0.757, 0.103)
Campaign information: ANES	-0.274	(-0.786, 0.237)
Social media: Pew	-0.315	(-0.663, 0.033)

Notes: Table shows the value of $\left(\hat{\beta}_c - \hat{\beta}\right)/\hat{\beta}_c$, where $\hat{\beta}_c$ and $\hat{\beta}$ are OLS estimates from a modified version of equation [2] in which s_t^g is replaced with $(\lambda_t^g + s_t^g)$ for λ_t^g a known scalar and the model is estimated with and without the constraint that $\rho = 0$, respectively. The equation is estimated on data from the 1996 to 2016 presidential election years and uses the 18–39, 40–64, and 65+ age groups. Each row shows the results for a separate internet use variable s_t^g, which measures the proportion of respondents in the age group that either use the internet (ANES and Pew Research Center), obtain campaign information online (ANES), or use social media (Pew Research Center). For the heterogeneous impact models, we set $\lambda_t^{18-39} = 0$, $\lambda_t^{40-64} = 0.1\left(s_t^{40-64}\right)$, and $\lambda_t^{65+} = 0.2\left(s_t^{65+}\right)$. For the social spillover models, we set $\lambda_t^{18-39} = 0$ and $\lambda_t^{40-65} = \lambda_t^{65+} = 0.1\left(s_t^{18-39}\right)$. Thus, the heterogeneous impact model assumes that the linear impact of the internet on polarization is 10 percent greater for those in the 40–64 age group, and 20 percent greater for those in the 65+ age group, than it is for those in the 18–39 age group. The social spillover model assumes that the internet participation of those in the 18–39 age group impacts the polarization of those in the other two age groups by an amount equal to 10 percent of the linear effect of the group's own internet participation. The 95 percent confidence intervals are constructed using the standard errors from a nonparametric bootstrap at the respondent level with 100 replicates. See section 3 of the SI appendix for details on the bootstrap procedure.

SI Table S5: Predicted internet, 1996

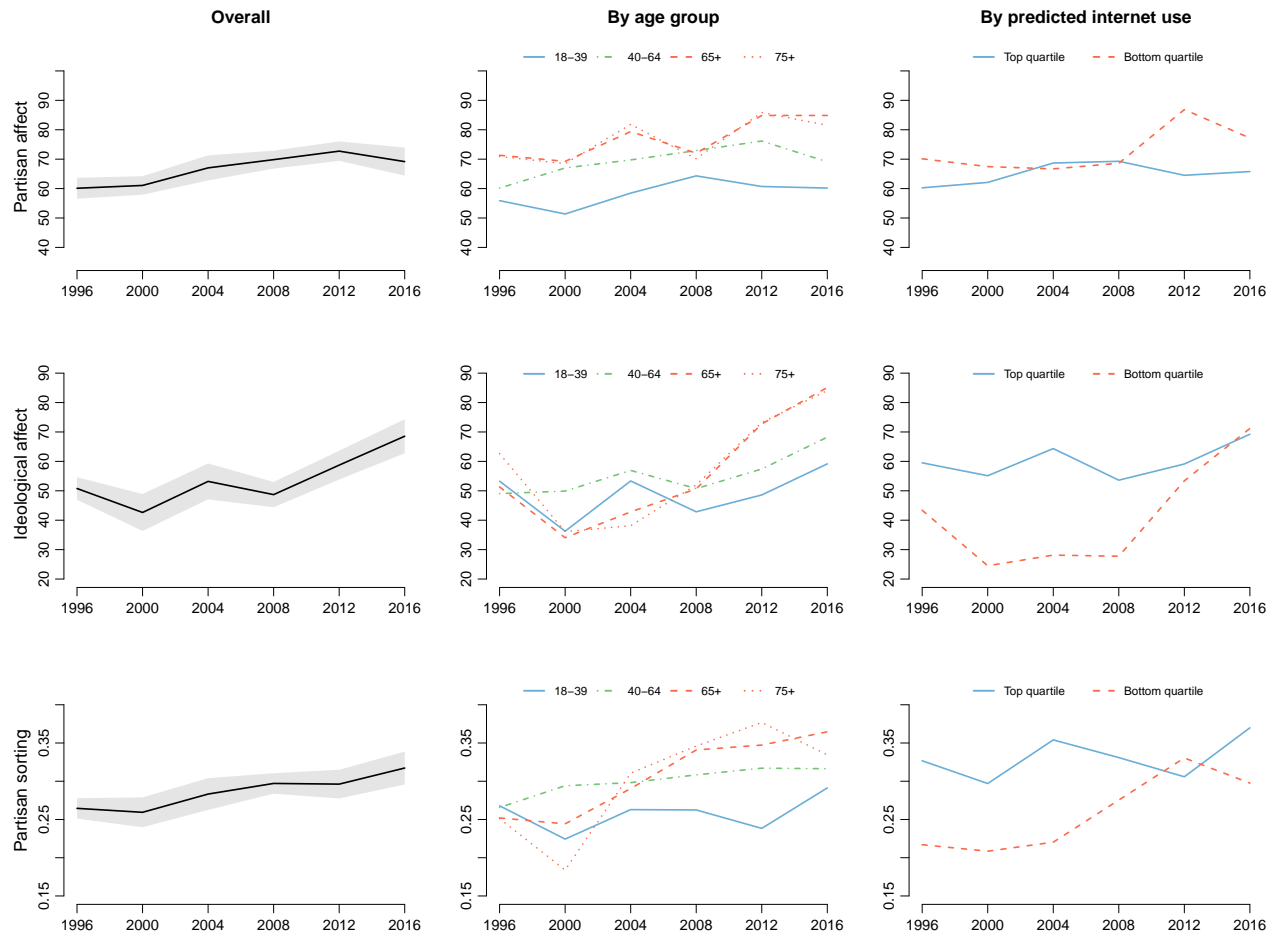|	Estimator: Weighted least squares
	Dependent variable: Internet use
Intercept	0.420
	(0.043)
Age Group: 40-64	-0.124
	(0.022)
Age Group: 65+	-0.269
	(0.030)
Gender: Male	0.008
	(0.020)
Race: Hispanic	0.046
	(0.045)
Race: Other	0.108
	(0.069)
Race: White	0.156
	(0.034)
Education: Grade School	-0.363
	(0.053)
Education: High School	-0.371
	(0.026)
Education: Some College	-0.146
	(0.029)
Region: South	0.081
	(0.022)
N	1513
R^2	0.220

Notes: Table shows the coefficients from a weighted least squares regression. Weights are the ANES survey weights. For estimation, the sample is restricted to respondents in 1996 who have valid responses to the questions needed to construct each independent and dependent variable. For prediction, the sample is restricted to respondents who have valid responses to the questions needed to construct each independent variable. Dependent variable is an indicator for whether an individual uses the internet taken from the ANES (see section 1 of the SI appendix for details on the variable construction). All covariates are indicator variables. Age groups are defined using *VCF0101* (for 1996–2012) and *V161267* (for 2016), gender is defined using *VCF0104* (for 1996–2012) and *V161267* (for 2016), race is defined using *VCF0105b* (for 1996–2012) and *V161310x* (for 2016), education is defined using *VCF0110* (for 1996–2012) and *V161270* (for 2016), and region is defined using *VCF0113* (for 1996–2012) and *V161010d* (for 2016). We treat the response '95. Other SPECIFY' as an invalid response to *V161270*. Standard errors are in parentheses.

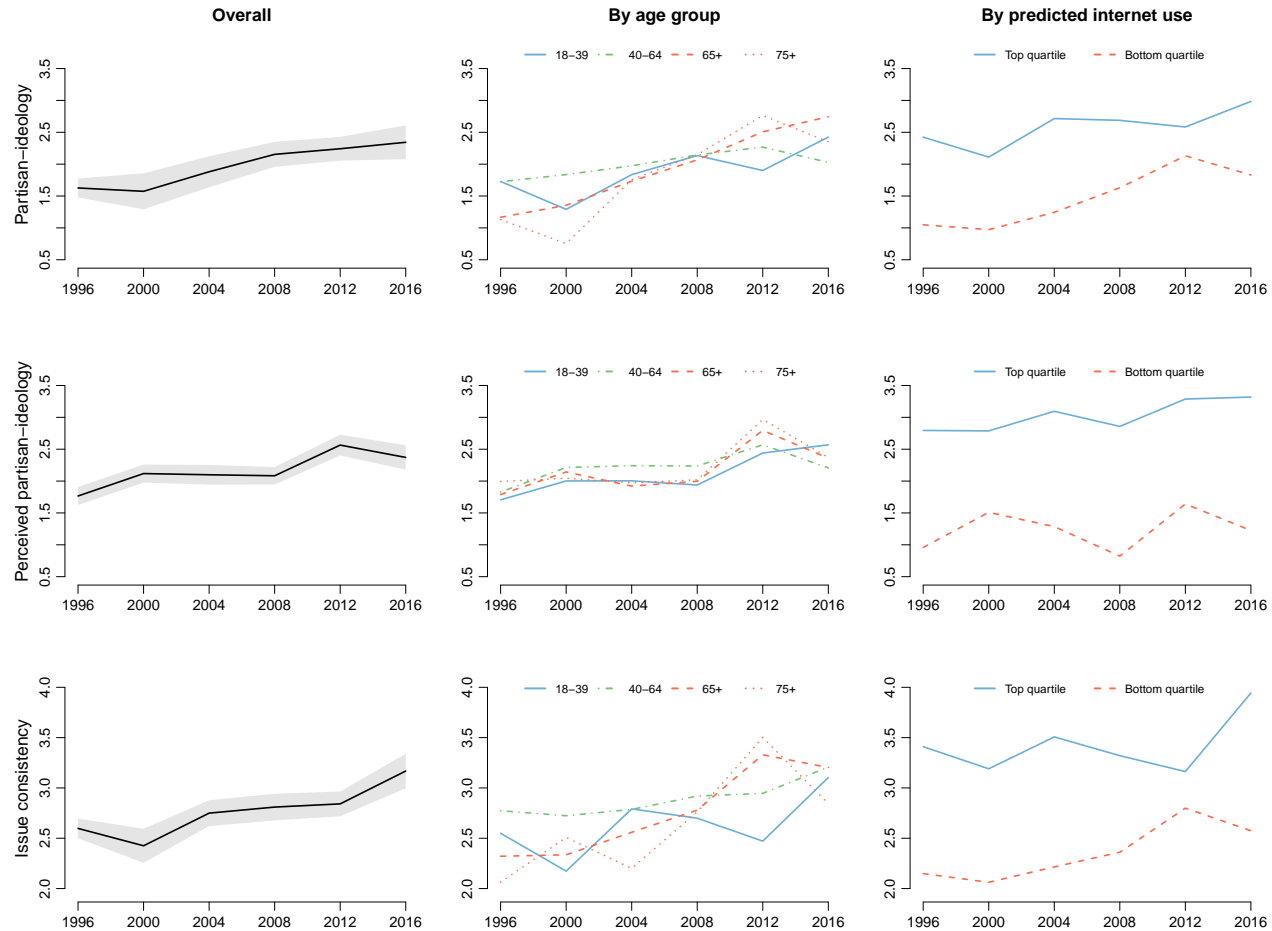SI Figure S1: Trends in internet use by age group, Pew data
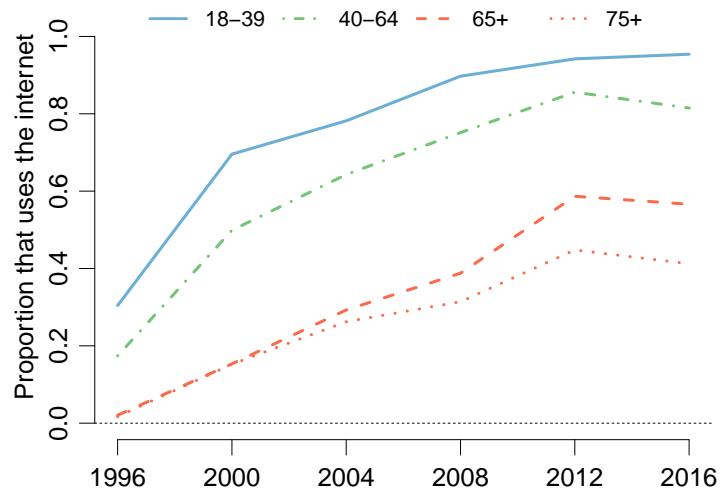
Notes: Plot shows trends in the weighted proportion of respondents that use the internet by age group according to the Pew Research Center surveys. See section 1 of the SI appendix for more details on the variable construction.

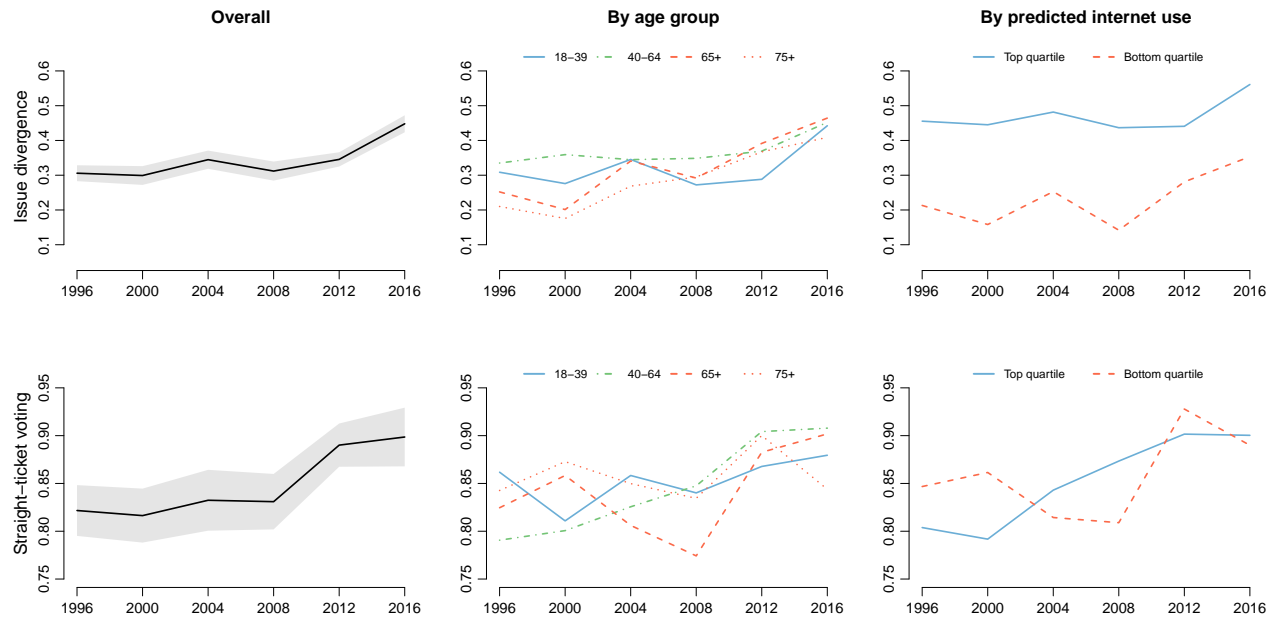SI Figure S2: Trends in polarization by demographic group, individual measures

Notes: Each row of plots displays the trends for a given polarization measure for various groups of respondents. The first figure in each row shows the overall measure with a pointwise 95 percent confidence interval constructed using a nonparametric bootstrap with 100 replicates. The second figure in each row shows the measure by age group. The third figure in each row shows the measure by quantiles of predicted internet use. See main text for definitions and for the construction of predicted internet use.

SI Figure S2: Trends in polarization by demographic group, individual measures (cont.)

Notes: Each row of plots displays the trends for a given polarization measure for various groups of respondents. The first figure in each row shows the overall measure with a pointwise 95 percent confidence interval constructed using a nonparametric bootstrap with 100 replicates. The second figure in each row shows the measure by age group. The third figure in each row shows the measure by quantiles of predicted internet use. See main text for definitions and for the construction of predicted internet use.

SI Figure S2: Trends in polarization by demographic group, individual measures (cont.)

Notes: Each row of plots displays the trends for a given polarization measure for various groups of respondents. The first figure in each row shows the overall measure with a pointwise 95 percent confidence interval constructed using a nonparametric bootstrap with 100 replicates. The second figure in each row shows the measure by age group. The third figure in each row shows the measure by quantiles of predicted internet use. See main text for definitions and for the construction of predicted internet use.

SI Figure S3: Trends in polarization by birth cohort
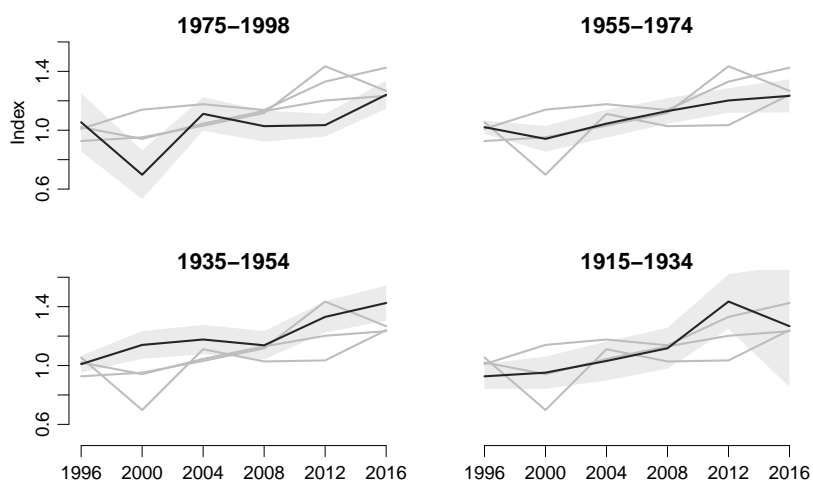
Notes: Each plot shows the polarization index for each of four birth cohorts. Each plot highlights the series for one cohort in bold. Shaded regions represent 95 percent pointwise confidence intervals for the bolded series constructed from a nonparametric bootstrap with 100 replicates. See main text for definitions and section 3 of the SI appendix for details on the bootstrap procedure.

SI Figure S4: Trends in polarization by age group and gender

Panel A: Males
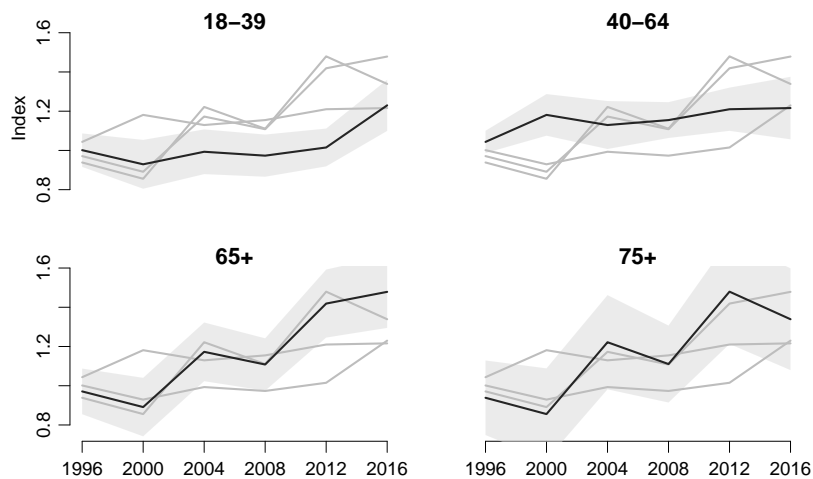
Panel B: Females
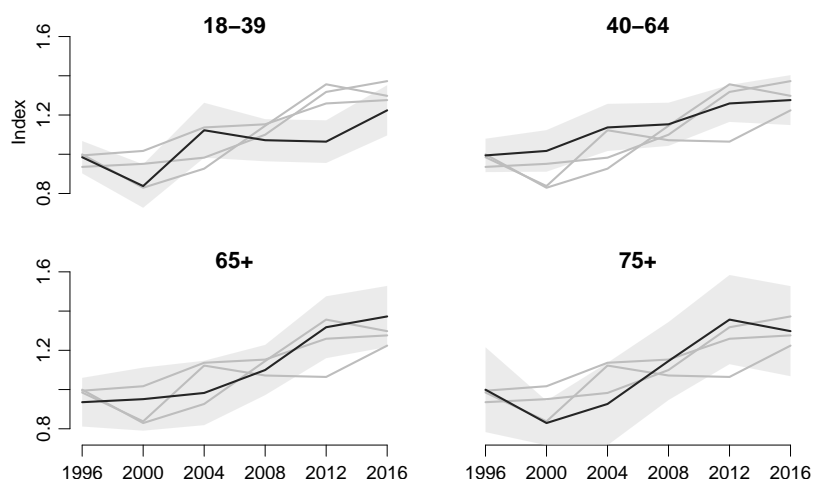
Notes: Each plot shows the polarization index for each of four age groups after restricting to respondents from a single gender. Panel A shows the plots for males. Panel B shows the plots for females. Each plot highlights the series for one age group in bold. Shaded regions represent 95 percent pointwise confidence intervals for the bolded series constructed from a nonparametric bootstrap with 100 replicates. See main text for definitions and section 1 of the SI appendix for details on the bootstrap procedure.

SI Figure S5: Trends in polarization by age group for the politically engaged

Panel A: Partisans
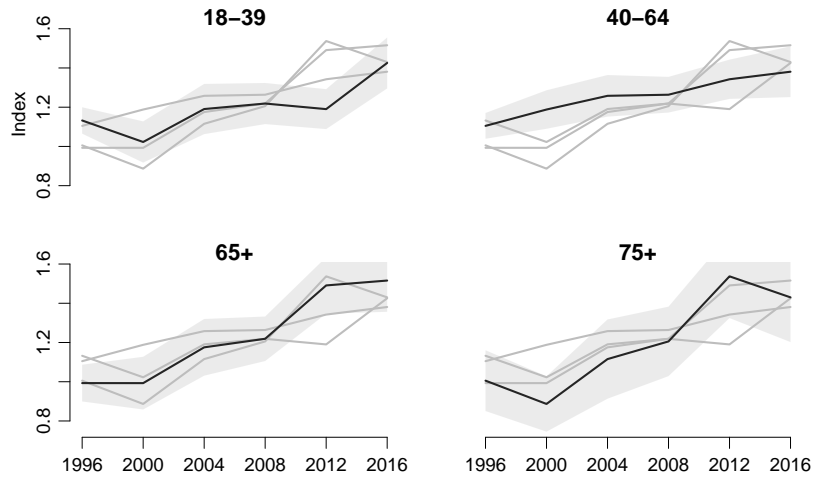

Panel B: Interested in election
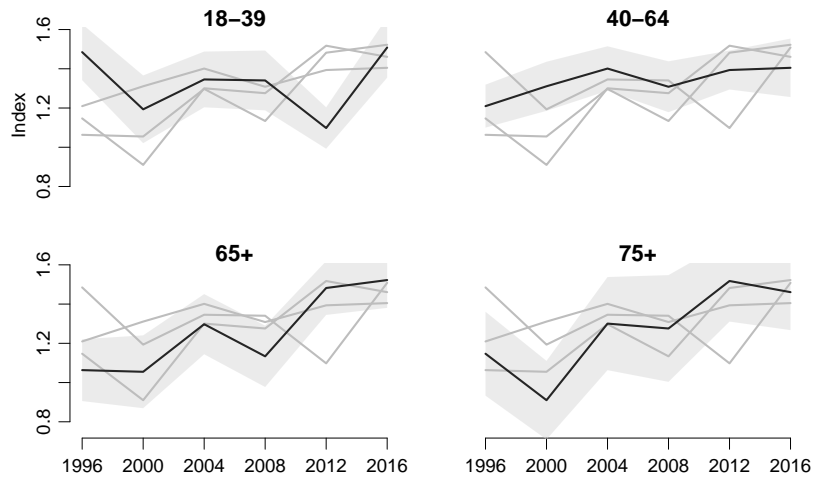
Notes: Each plot shows the polarization index for each of four age groups after restricting to respondents with positive responses to a proxy for political engagement. Panel A shows the plots for respondents who self-identify with either the Republican or Democratic party (excluding leaners). Panel B shows the plots for respondents who self-identify as being "very much interested" in the upcoming election. Each plot highlights the series for one age group in bold. Shaded regions represent 95 percent pointwise confidence intervals for the bolded series constructed from a nonparametric bootstrap with 100 replicates. See main text for definitions and section 3 of the SI appendix for details on the bootstrap procedure.

SI Figure S6: Trends in polarization by age group, extended series
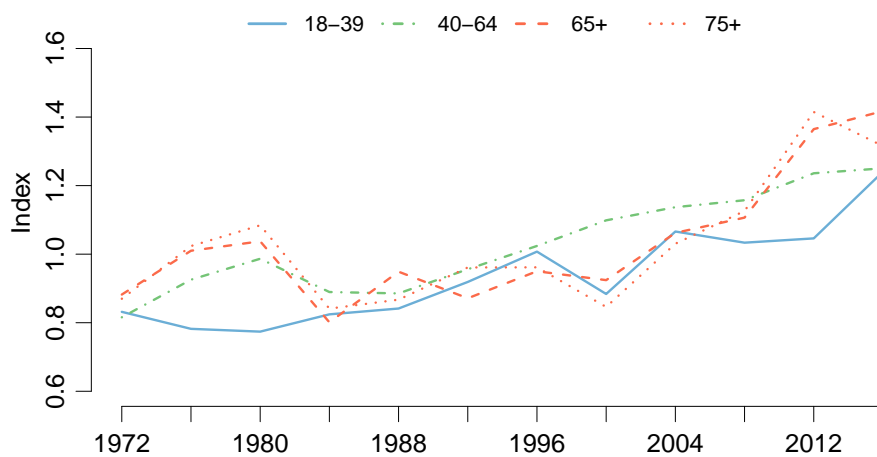
Notes: Plot shows the polarization index by age group from 1972 to 2016. After normalizing each measure by dividing by the 1996 value of the measure on the full sample, we take the average across all polarization measures available in a given year for a given age group.

15

SI Figure S7: Linear model of polarization, internet use (ANES)
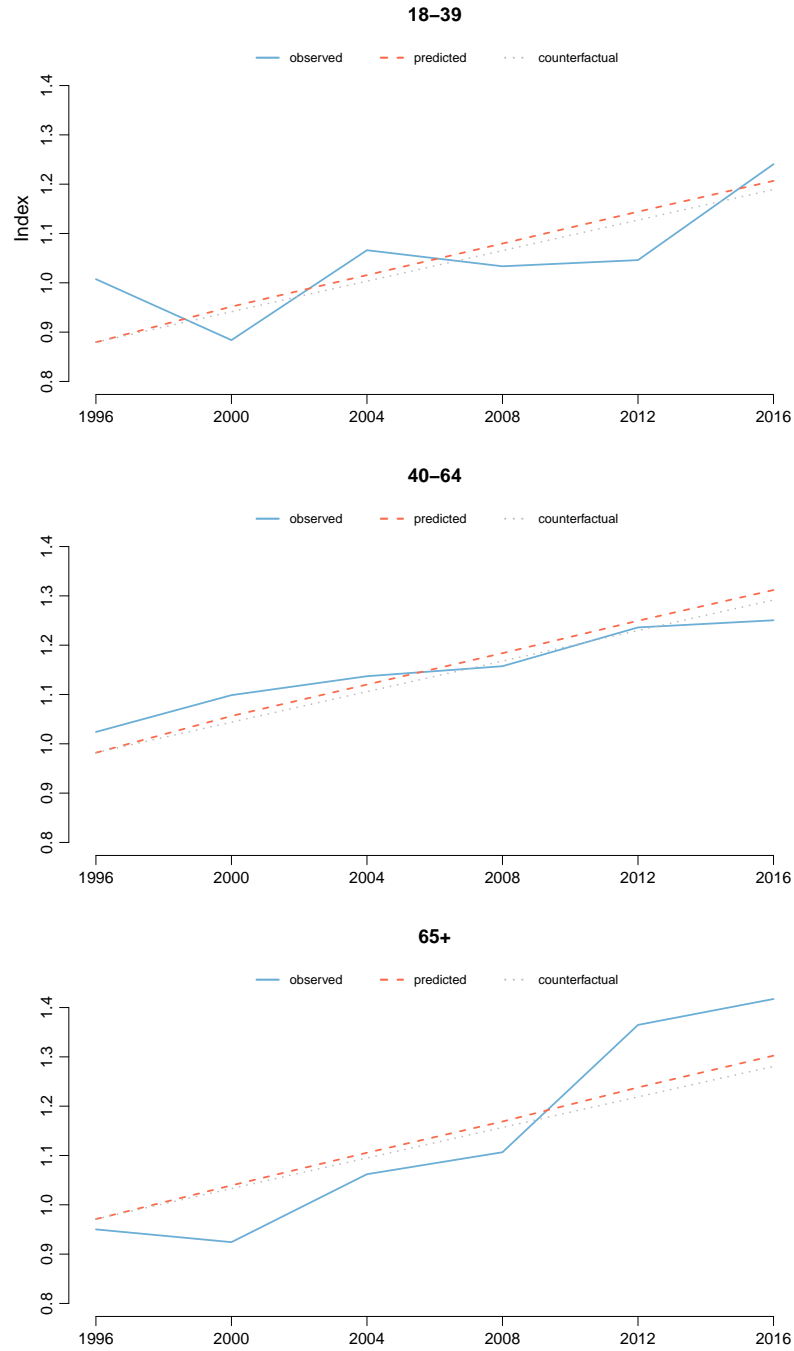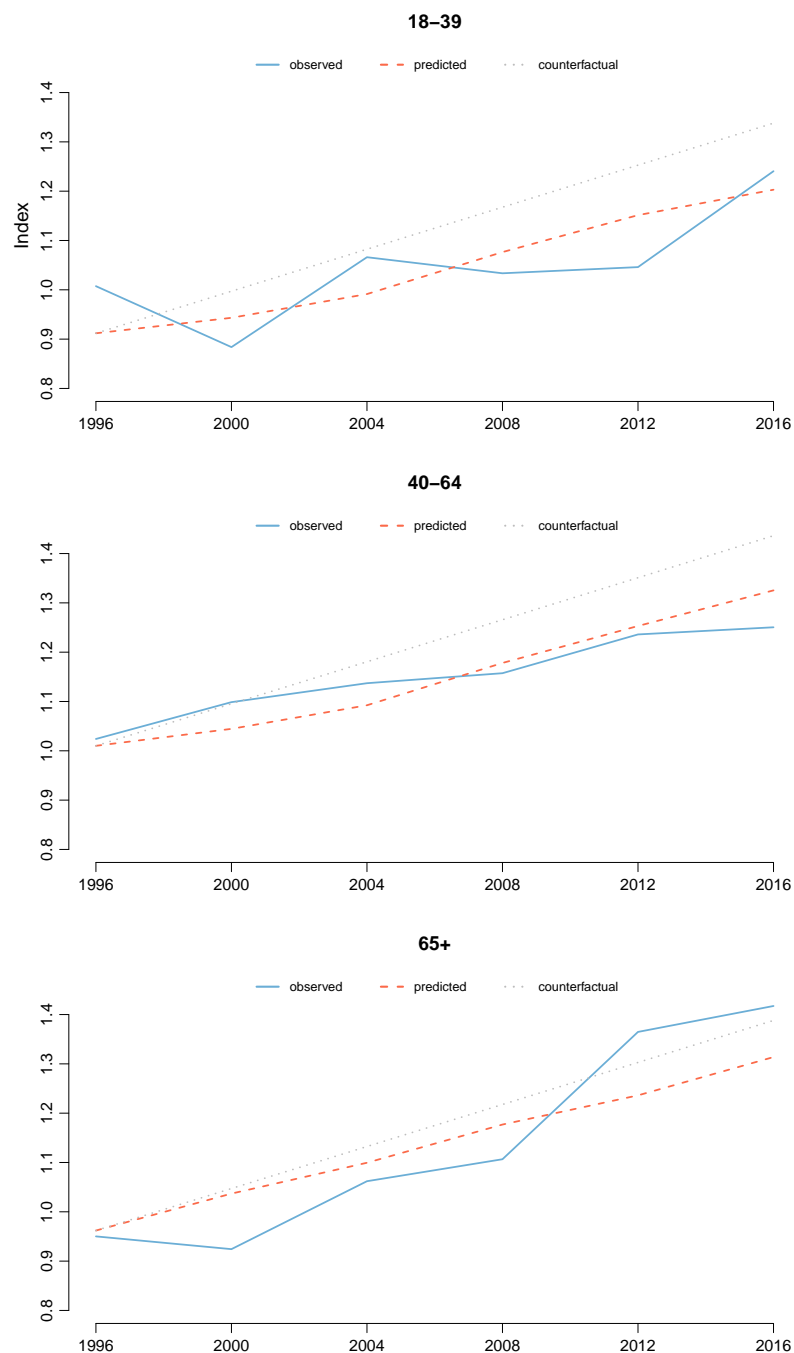
18–39

40–64

65+

Notes: Each plot shows the observed, the predicted, and a counterfactual series for a given age group under a linear model of the internet's impact on polarization using the proportion s_t^g of the group that uses the internet according to the ANES as the internet variable of interest. The series labeled "observed" is the observed polarization index M_t^g for each age group g in each presidential election year t. The series labeled "predicted" is the predicted polarization from an OLS estimate of the model in equation [2], as displayed in column 2 of SI appendix table S3. The series labeled "counterfactual" shows the predicted polarization from the same OLS estimate when we suppose that the proportion s_t^g of the group that uses the internet remains constant at its 1996 level.

16

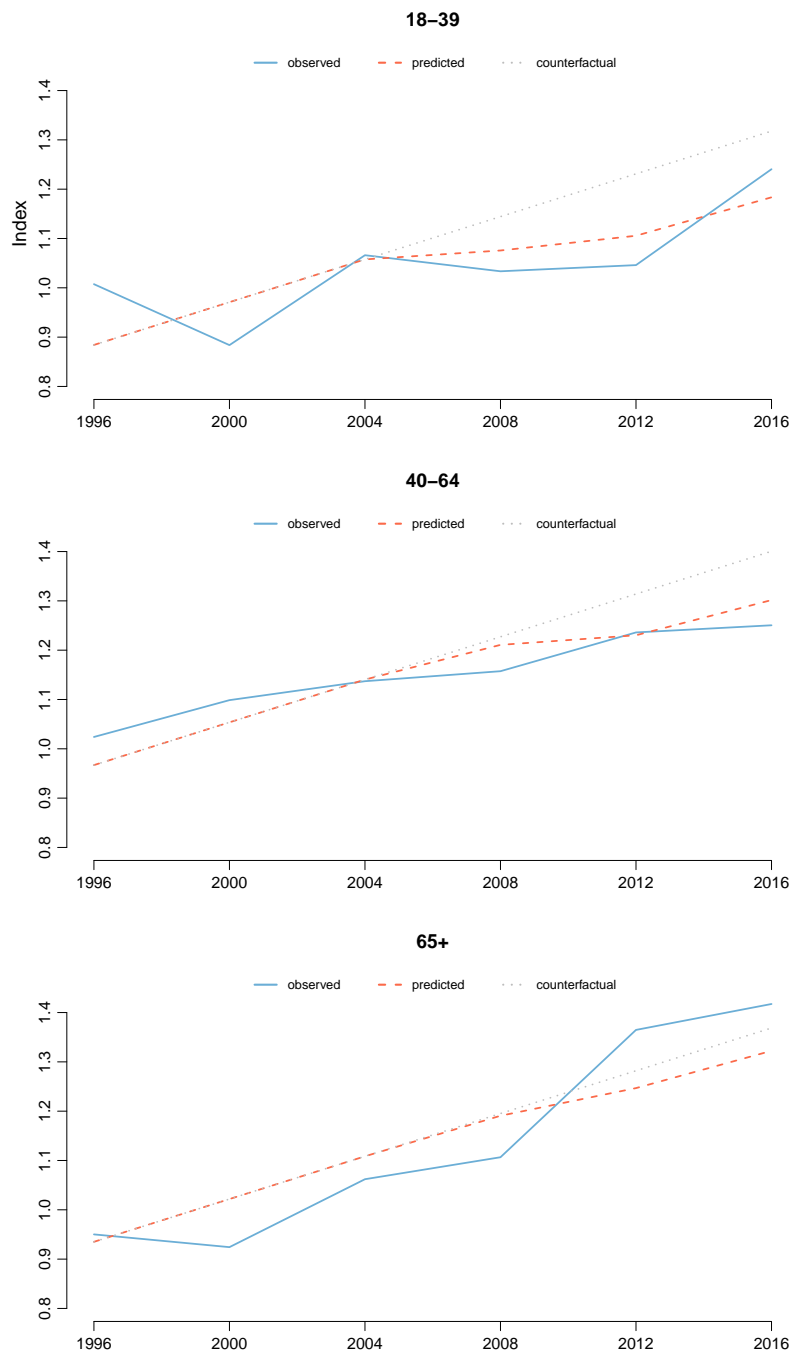SI Figure S8: Linear model of polarization, internet use (Pew)
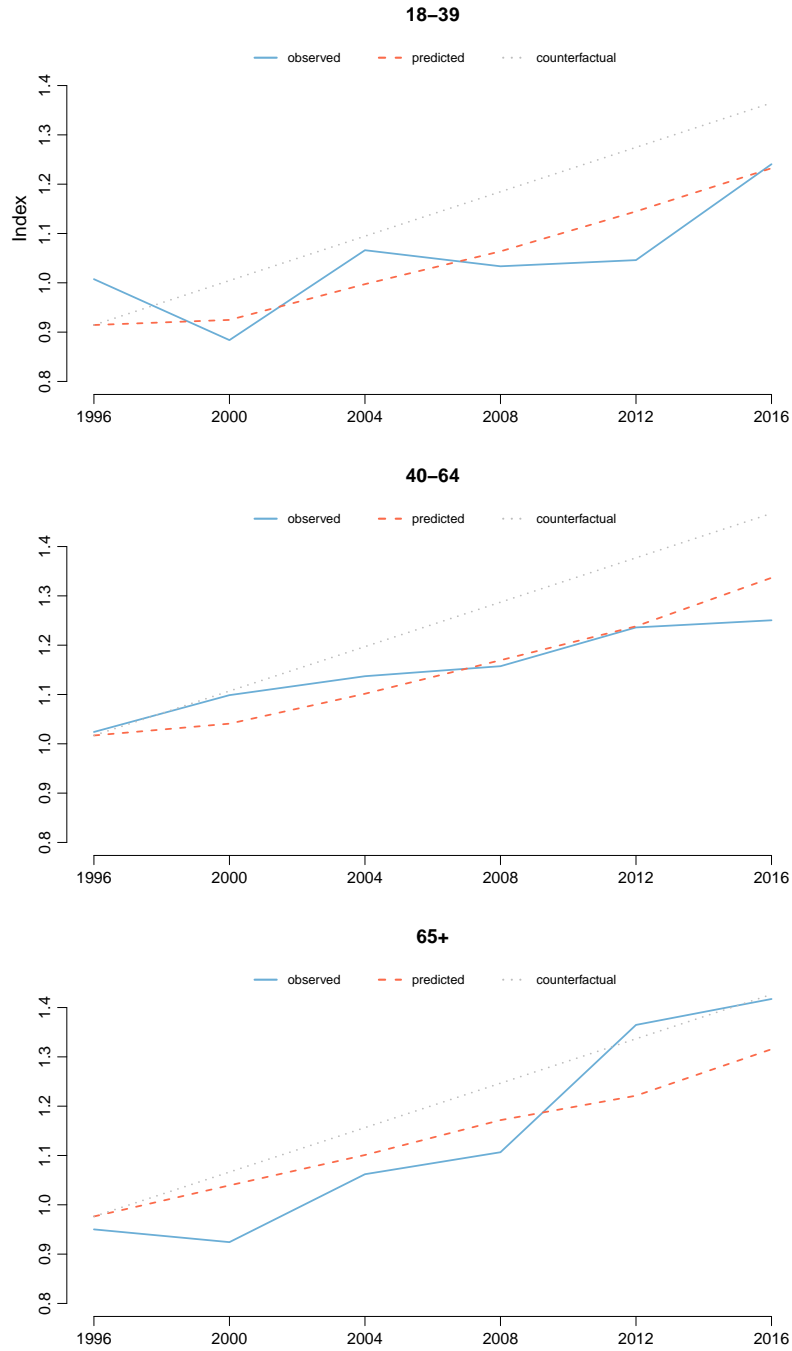
Notes: Each plot shows the observed, the predicted, and a counterfactual series for a given age group under a linear model of the internet's impact on polarization using the proportion s_t^g of the group that uses the internet according to the Pew Research Center as the internet variable of interest. The series labeled "observed" is the observed polarization index M_t^g for each age group g in each presidential election year t. The series labeled "predicted" is the predicted polarization from an OLS estimate of the model in equation [2], as displayed in column 3 of SI appendix table S3. The series labeled "counterfactual" shows the predicted polarization from the same OLS estimate when we suppose that the proportion s_t^g of the group that uses the internet remains constant at its 1996 level.

SI Figure S9: Linear model of polarization, campaign information (ANES)


18–39

40–64

65+

Notes: Each plot shows the observed, the predicted, and a counterfactual series for a given age group under a linear model of the internet's impact on polarization using the proportion s_t^g of the group that obtains campaign information online according to the ANES as the internet variable of interest. The series labeled "observed" is the observed polarization index M_t^g for each age group g in each presidential election year t. The series labeled "predicted" is the predicted polarization from an OLS estimate of the model in equation [2], as displayed in column 4 of SI appendix table S3. The series labeled "counterfactual" shows the predicted polarization from the same OLS estimate when we suppose that the proportion s_t^g of the group that obtained campaign information online remains constant at its 1996 level.

SI Figure S10: Linear model of polarization, social media use (Pew)

18–39

40–64

65+

Notes: Each plot shows the observed, the predicted, and a counterfactual series for a given age group under a linear model of the internet's impact on polarization using the proportion s_t^g of the group that uses social media according to the Pew Research Center as the internet variable of interest. The series labeled "observed" is the observed polarization index M_t^g for each age group g in each presidential election year t. The series labeled "predicted" is the predicted polarization from an OLS estimate of the model in equation [2], as displayed in column 5 of SI appendix table S3. The series labeled "counterfactual" shows the predicted polarization from the same OLS estimate when we suppose that the proportion s_t^g of the group using social media remains constant at its 1996 level.

SI Figure S11: Trends in the first principal component of polarization measures by age group
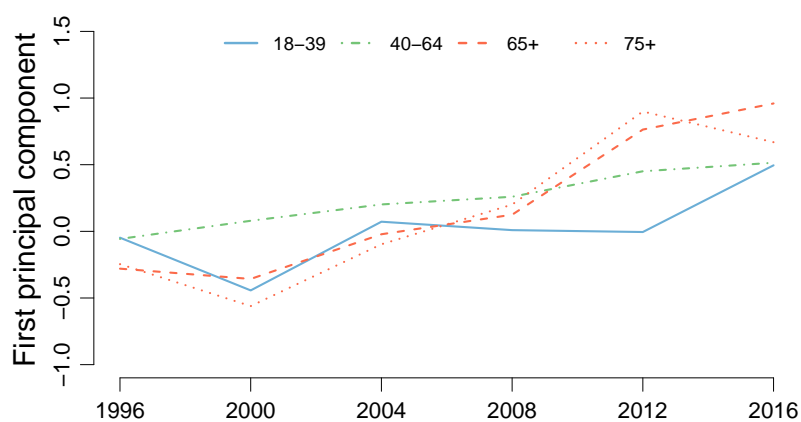
Notes: Plot shows the predicted first principal component of the polarization measures by age group between 1996 and 2016. The linear coefficients of the first principal component of the normalized polarization measures m_t/m_{1996} are estimated using data for presidential election years between 1984 and 2016. These linear coefficients are then used to predict the trend in the first principal component for each age group. The correlation between the first principal component of the polarization measures and the index M_t between 1984 and 2016 is over .999.

SI Figure S12: Trends in polarization by predicted internet use, using 1996 data to predict internet use
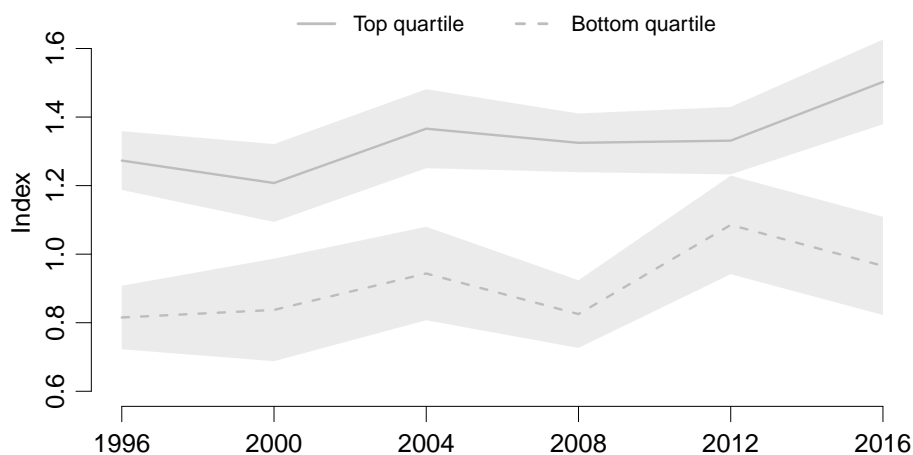
Notes: Plot shows the polarization index broken out by quartile of predicted internet use within the given survey year, where predicted internet use is based on the weighted least squares regression reported in SI appendix table S5. Respondents with invalid or missing covariates are excluded. The bottom quartile includes values that are at or below the 25th percentile, while the top quartile includes values greater than the 75th percentile. For a given measure and group, the polarization value is normalized by the 1996 value of the polarization measure in the full sample. The index is then the average of these normalized polarization measures for each group.

1 Data Construction Details

The list below gives details on the construction of some of the variables used. All variables whose name begins with "VCF" come from the 1948–2012 Time Series Cumulative Data File. There was no cumulative file that included 2016 at the time of analysis, so we use separate variable identifiers for 2016 respondents. All information on survey questions in the ANES is taken from the survey codebooks.

1. Internet use (ANES): Our measure of internet use comes from:

   - 1996–2008 (*VCF0744* in the 1948–2012 Time Series Cumulative file): "Do you have access to the Internet or the World Wide Web [exc. 2008: ('the Web')]?"
   - 2012 (*prmedia_useinet* in the 2012 Time Series Study): "Do you or anyone in this household use the Internet at any location?"
   - 2016 (*V161007* in the 2016 Time Series Study): See 2012 wording.

2. Campaign information online (ANES): Our measure of obtaining campaign information online comes from:

   - 1996–2004 (*VCF0745* in the 1948–2012 Time Series Cumulative file): "Have you seen any information about this election campaign on (the Internet/the Web)?"
   - 2008 (*V085010* and *V085021* in the 2008 Time Series Study): "Did you read, watch, or listen to any information about the campaign for President on the Internet?"[1]
   - 2012 (*mediapo_net* in the 2012 Time Series Study): See 2008 wording.
   - 2016 (*V161363d* in the 2016 Time Series Study): An indicator for whether respondents "heard anything about the presidential campaign" on "Internet sites, chat rooms, or blogs."

3. Internet use (Pew): We use one Pew Research Center survey from each presidential election year between 1996 and 2016. (See references in main text for exact surveys.)

   - 1996: "Do you ever use a computer at work, school or home to connect with computer bulletin boards, information services such as America Online or Prodigy, or other computers over the Internet?"
   - 2000 and 2004: "Do you ever go online to access the Internet or World Wide Web or to send and receive email?"
   - 2008 and 2012: Positive responses to at least one of:
     - "Do you use the internet, at least occasionally?"
     - "Do you send or receive email, at least occasionally?"
   - 2016: "Do you use the internet or email, at least occasionally?"

---

[1]In 2008, ANES randomly split respondents across different questionaires. As far as we can tell, the wording and responses for the campaign information question are the same, but the other questions in close proximity are different.

Surveys in 2012 and 2016 also ask respondents "Do you access the internet on a cell phone, tablet or other mobile handheld device, at least occasionally?" We ignore this question when creating our internet use variable for consistency across years. In 1996 and 2000, respondents are only asked the internet use question if they reported using a computer; we treat a negative answer to the computer use question as a negative answer to the internet use question and a missing answer to the computer use question as a missing answer to the internet use question.

4. Social media use (Pew): In addition to the 2008, 2012, and 2016 Pew Research Center surveys mentioned in the definition of the Pew Research Center internet use variable above, we use Pew Research Center surveys from 2005 and 2011 to construct the social media use variable. (See references in main text for exact surveys.) These surveys ask respondents who claim to (see exact wording above) "use the internet" or "send or receive email" (and, in 2012 and 2016, those who "access the internet on a cell phone, tablet or other mobile handheld device") at least occasionally:

   - 1996: "...Do you ever... [u]se online social or professional networking sites like Friendster or LinkedIn"
   - 2008: "...Do you ever use the internet to... [u]se a social networking site like MySpace, Facebook, or LinkedIn.com"
   - 2011: Positive responses to at least one of: "...Do you ever use the internet to..."
     - "Use a social networking site like MySpace, Facebook or LinkedIn"
     - "Use Twitter"
   - 2012: Positive responses to at least one of: "...Do you ever..."
     - "Use a social networking site like Facebook, LinkedIn or Google Plus"
     - "Use Twitter"
   - 2016: Postive responses to at least one of: "...Do you ever use the internet or a mobile app to..."
     - "Use Twitter"
     - "Use Instagram"
     - "Use Pinterest"
     - "Use Facebook"
     - "Use LinkedIn"

   The survey in 2016 also asks respondents whether they use "messaging apps," "an app that automatically deletes the message you send," or "social media apps like YikYak, Whisper, After School, or Rumr" which we exclude. For all surveys, we treat refusals or "Don't know" responses to the social network or any of its prerequisite questions as missing observations.

5. Age: For pre-2016 respondents, we use *VCF0101* (2012 wording: "What is the month, day and year of your birth?"). This value is then converted into an age and top-coded by the ANES. The top-coding varies across years, but is at least 90 and thus does not impact our age group assignment. However, it does impact our cohort assignment and we treat

the coded value as-is. A few respondents are 17 years old at the time of the survey, but will be 18 at the time of the election, and are treated as 18-year-olds. We use variable *V161267* for 2016 respondents.

6. Party (P_i) and ideology (B_i): For pre-2016 respondents, party and ideology affiliation come from questions *VCF0301* ("Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent, or what? [If R/D] Would you call yourself a strong [R/D] or a not very strong [R/D]? [If Independent, Other, or No Preference] Do you think of yourself as closer to the Republican or Democratic party?") and *VCF0803* ("We hear a lot of talk these days about liberals and conservatives. Here is a 7-point scale on which the political views that people might hold are arranged from extremely liberal to extremely conservative. Where would you place yourself on this scale, or haven't you thought much about this?") respectively, and are shifted downwards by four in order to range from -3 to 3. Individuals who initially report not being affiliated with either the Republican or Democratic parties, but report being closer to one party than the other, are considered "leaners" (values of -1 or 1). Individuals that place themselves exactly in the middle of the 7-point ideology scale are considered "strict moderates" (value of 0). Also, as mentioned in the main draft, individuals who respond "Don't know" or "Haven't though much about it" to the question on self-reported ideology are treated as having a missing or non-valid response. For 2016 respondents, we use *V161158x* and *V161126* after performing the same manipulations.

7. Survey weights (w_i): For the ANES data, we use the type-0, face-to-face survey weights from variable *VCF0009x* for respondents in 1996–2012, and the post-election, face-to-face survey weights from *V160102f* for 2016 respondents. For the Pew Research Center data, we use the survey weights included with each survey.

8. Affect (A_i): We use the ANES thermometer questions that ask respondents to rate various groups from 0–100, where higher values indicate more favorable feelings and a value of 50 designates neutral feelings. The exact wording of the questions has changed over time. We use responses for the following groups: liberals (*VCF0211* before 2016 and *V162097* in 2016), conservatives (*VCF0212* before 2016 and *V162101* in 2016), Democratic Party (*VCF0218* before 2016 and *V161095* in 2016), and Republican Party (*VCF0224* before 2016 and *V161096* in 2016). Responses between 97 and 100 are top-coded by the ANES at 97 before 2016 and are not top-coded in 2016. The 2012 preface for these questions is as follows: "Please look at page [preload: prepg_c] of the booklet. I'd like to get your feelings toward some of our political leaders and other people who are in the news these days. I'll read the name of a person and I'd like you to rate that person using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favorable and warm toward the person. Ratings between 0 degrees and 50 degrees mean that you don't feel favorable toward the person and that you don't care too much for that person. You would rate the person at the 50 degree mark if you don't feel particularly warm or cold toward the person. If we come to a person whose name you don't recognize, you don't need to rate that person. Just tell me and we'll move on to the next one."

9. Perceived ideology of parties ($\tilde{R}, \tilde{D}$): For pre-2016 respondents, we use questions *VCF0503* and *VCF0504*, and we shift responses down by four. For 2016 respondents, we use

V161130 and *V161131*, and we perform the same downward shift. The wording for 2012 is as follows: "We hear a lot of talk these days about liberals and conservatives. Here is a seven-point scale on which the political views that people might hold are arranged from extremely liberal to extremely conservative.... Where would you place the Republican party on this scale?"

10. Issue questions: The seven questions used for the issue consistency and issue divergence measures are as follows with 2012 wording:

   - 7-point ideological position (*VCF0803* before 2016 and *V161126* in 2016)

   - Aid to blacks (*VCF0830* before 2016 and *V161198* in 2016)
     - "Some people feel that the government in Washington should make every effort to improve the social and economic position of blacks. Suppose these people are at one end of a scale, at point 1. Others feel that the government should not make any special effort to help blacks because they should help themselves. Suppose these people are at the other end, at point 7. And, of course, some other people have opinions somewhere in between, at points 2, 3, 4, 5, or 6. Where would you place yourself on this scale, or haven't you thought much about this?"

   - Foreign defense spending (*VCF0843* before 2016 and *V161181* in 2016)
     - "Some people believe that we should spend much less money for defense.... Others feel that defense spending should be greatly increased...."

   - Government responsibility to guarantee jobs and income (*VCF0809* before 2016 and *V161189* in 2016)
     - "Some people feel the government in Washington should see to it that every person has a job and a good standard of living.... Others think the government should just let each person get ahead on their own...."

   - Government health insurance plan (*VCF0806* before 2016 and *V161184* in 2016)
     - "There is much concern about the rapid rise in medical and hospital costs. Some people feel there should be a government insurance plan which would cover all medical and hospital expenses for everyone.... Others feel that all medical expenses should be paid by individuals through private insurance plans like Blue Cross or other company paid plans...."

   - Government services and spending (*VCF0839* before 2016 and *V161178* in 2016)
     - "Some people think the government should provide fewer services even in areas such as health and education in order to reduce spending.... Other people feel it is important for the government to provide many more services even if it means an increase in spending...."

   - Abortion legislation (*VCF0838* before 2016 and *V161232* in 2016)
     - "There has been some discussion about abortion during recent years. Which one of the opinions on this page best agrees with your view? You can just tell me the number of the opinion you choose. 1. By law, abortion should never

be permitted. 2. The law should permit abortion only in case of rape, incest, or when the woman's life is in danger. 3. The law should permit abortion for reasons other than rape, incest, or danger to the woman's life, but only after the need for the abortion has been clearly established. 4. By law, a woman should always be able to obtain an abortion as a matter of personal choice."

All questions are answered on a 7-point, liberal-to-conservative scale (except government services and spending whose scale is conservative-to-liberal of the same order and abortion legislation which has four available positions). Liberal positions (points 1–3 on the scale) are coded as -1, moderate positions (point 4 on the scale) and refusals (e.g., "Don't know" or "Haven't thought much about it") are coded as 0, and conservative positions (points 5–7 on the scale) are coded as 1. For the abortion questions, we code conservative positions (positions 1 and 2) as 1, moderate positions (position 3) and refusals (e.g., "Don't know" or "Haven't thought much about it") as 0, and liberal positions (position 4) as -1.

11. Straight-ticket voting: For pre-2016 respondents, we use *VCF0709*, which aggregates responses from previous questions regarding for whom the respondent voted in the most recent House and Presidential elections. For 2016 respondents, we use *V162040*, *V162041*, and *V162058x* to construct the same measure.

2   Measure Details

The list below gives the details on the construction of the polarization measures and a specific reference to where they are used elsewhere in the literature.

Let S_t denote the set of all face-to-face survey respondents in year t who have a valid response for their age. Let P_i denote a 7-point measure of individual i's party from strong Democrat (coded as -3) to Independent (coded as 0) to strong Republican (coded as 3), and B_i denote a 7-point measure of individual i's ideological affiliation from strong liberal (coded as -3) to moderate (coded as 0) to strong conservative (coded as 3). Throughout, we let $R_t := \{i : P_i > 1\}$ and $R_t^+ := \{i : P_i > 0\}$ denote the sets of Republicans excluding and including leaners respectively in time t. We define D_t and D_t^+ analogously. Likewise, we let $C_t := \{i : B_i > 0\}$ and $L_t := \{i : B_i < 0\}$ denote the sets of conservatives and liberals, respectively. We let w_i denote the ANES survey weights for respondent i.

As noted in the main draft, in computing each polarization measure, we restrict the sample to respondents with valid, non-missing responses (as defined by the ANES codebooks) to each of the relevant questions used in constructing the measures. We do not explicitly adjust our set notation to account for this restriction.

1. Partisan affect polarization: See "Affective Distance" from table A1 in the online appendix of (1) and figure 9 in (2). Formally, partisan affect polarization is

$$\frac{1}{\sum_{i \in D_t^+} w_i} \sum_{i \in D_t^+} w_i(A_i^D - A_i^R) + \frac{1}{\sum_{i \in R_t^+} w_i} \sum_{i \in R_t^+} w_i(A_i^R - A_i^D),$$

where A_i^R and A_i^D denote individual i's feelings towards the Republican and Democratic parties, respectively, on a 0–100 scale, with higher values representing more favorable feelings.

2. Ideological affect polarization: See "Affective Distance" from table A2 in the online appendix of (1). Formally, ideological affect polarization is

$$\frac{1}{\sum_{i \in L_t} w_i} \sum_{i \in L_t} w_i(A_i^L - A_i^C) + \frac{1}{\sum_{i \in C_t} w_i} \sum_{i \in C_t} w_i(A_i^C - A_i^L),$$

where A_i^C and A_i^L denote an individual i's feelings towards conservatives and liberals, respectively, on a 0–100 scale, with higher values representing more favorable feelings.

3. Partisan sorting: See figure 1 in (3). Details on the implementation are taken from (4). Formally, partisan sorting is

$$\frac{1}{\sum_{i \in S_t} w_i} \sum_{i \in S_t} \frac{w_i}{105}[g(|P_i - B_i| + 1)(|P_i| + 1)(|B_i| + 1) - 7],$$

where $g(x) = \max_{i \in \cup_t S_t}(|P_i - B_i| + 1) + \min_{i \in \cup_t S_t}(|P_i - B_i| + 1) - x$.

4. Partisan-ideology polarization: See pg. 547 of (5). Formally, partisan-ideology polarization is

$$\frac{1}{\sum_{i \in R_t} w_i} \sum_{i \in R_t} B_i w_i - \frac{1}{\sum_{i \in D_t} w_i} \sum_{i \in D_t} B_i w_i.$$

5. Perceived partisan-ideology polarization: See pg. 283 of (6). Formally, perceived partisan-ideology polarization is

$$\frac{1}{\sum_{i \in S_t} w_i} \sum_{i \in S_t} w_i(\tilde{R}_i - \tilde{D}_i)$$

where $\tilde{R}_i$ and $\tilde{D}_i$ denote individual i's perception on how conservative the Republican or Democratic party is, respectively, on the 7-point liberal-to-conservative scale.

6. Issue consistency: See pg. 544 of (5) where they outline their 7-point measure of ideological polarization. While (5) further collapse this 7-point scale to label individuals as low, moderate, or high and report the percentage of respondents with ideologically consistent views in their table 1, we simply report the average of the 7-point measure. Footnote 2 of (5) suggests that "using a simple additive index consisting of the same items produces nearly identical results concerning the trend in polarization between 1984 and 2004." Formally, issue consistency is

$$\frac{1}{\sum_{i \in S_t} w_i} \sum_{i \in S_t} w_i |\sum_{k \in K} k_i|,$$

where $k_i \in \{1, 0, -1\}$ denotes a conservative, moderate, or liberal response respectively to one of seven policy questions K. (See SI section 1 for details on the policy questions used.)

7. Issue divergence: See table 3 in (5). Each correlation is computed for the set of respondents with valid, non-missing responses to both the party identification question P_i and the given issue question $k \in K$. Formally, issue divergence is

$$\frac{1}{|K|} \sum_{k \in K} \mathrm{cor}_{i \in R_t^+ \cup D_t^+}(\mathbf{1}_{P_i > 0}, k_i),$$

where cor calculates Kendall's tau between the party indicator $\mathbf{1}_{P_i > 0}$ and the response k_i to a given issue question. We use the same set of policy questions to define issue divergence that we use to define issue consistency, which means our set of questions differs somewhat from that used in table 3 in (5).

8. Straight-ticket voting: See figure 3 in (7).

3   Bootstrap Details

When conducting inference using a bootstrap, we use a nonparametric bootstrap with 100 replicates and take the standard deviation across these replicates to form standard errors σ. Our 95 percent confidence intervals are then $(\hat{\mu} - 1.96\sigma, \hat{\mu} + 1.96\sigma)$ where $\hat{\mu}$ is our point estimate for a given statistic. For the ANES data, we draw replicates by sampling, with replacement, from the full sample of face-to-face survey respondents in year t who have a valid response for their age and construct measures of polarization for each replicate following the procedure outlined in the main text and SI appendix sections 1 and 2. For the Pew Research Center data, we draw replicates by sampling, with replacement, from the full sample of respondents in year t with valid responses for their age and compute the internet and social media use variables as described in the main text and SI appendix section 1. If a statistic depends on both the ANES and the Pew Research Center data, we bootstrap the data separately for each source before computing the statistic. Separately, for each year, demographic group, and polarization measure, we drop bootstrap replicates which have insufficient observations to construct a given polarization measure. For the index, we simply take the average across the normalized polarization measures with sufficient observations for a given bootstrap replicate and do not drop the replicate if some polarization measures have insufficient observations. For the bootstraps involving predicted internet use, we re-estimate equation [1] in the main text and re-assign quantiles for each replicate.

References

1. Iyengar, Shanto, Gaurav Sood, and Ypthach Lelkes. 2012. Affect, not ideology: A social identity perspective on polarization. *Public Opinion Quarterly*. 76(3): 405–431.

2. Gentzkow, Matthew. 2016. Polarization in 2016. *Toulouse Network for Information Technology Whitepaper.* Available at http://www.techpolicy.com/Blog/March-2016/Polarization-in-2016.aspx. Accessed July 20, 2017.

3. Mason, Lilliana. 2015. "I disrespectfully agree": The differential effects of partisan sorting on social and issue polarization. *American Journal of Political Science*. 59(10): 128–145.

4. Mason, Lilliana. 2013. Replication data for: "'I disrespectfully agree': The differential effects of partisan sorting on behavioral and issue polarization." Harvard Dataverse, V3. Available at https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/22346. Accessed January 20, 2017.

5. Abramowitz, Alan I. and Kyle L. Saunders. 2008. Is polarization a myth? *The Journal of Politics*. 70(2): 542–555.

6. Davis, Nicholas T. and Johanna L. Dunaway. 2016. Party polarization, media choice, and mass partisan-ideological sorting. *Public Opinion Quarterly*. 80(S1): 272–297.

7. Hetherington, Marc J. 2001. Resurgent mass partisanship: The role of elite polarization. *The American Political Science Review*. 95(3): 619–631.