

Online Appendix:
Let's Talk Politics
A Naive Approach for Measuring Political Sophistication

Patrick W. Kraft*

A Information on Open-Ended Responses and Discursive Sophistication	1
I Distribution of Word Counts in Open-Ended Responses	1
II Overview of Topic Proportions	2
III Discursive Sophistication Components	5
B Pre-Processing and Topic Model Specification	6
I PreText Analysis	6
II Robustness Checks for Varying Model Specifications	7
C Additional Information on Remaining Variables	9
I Item Selection and Recoding	9
II Variable Distributions – 2012 ANES	11
III Variable Distributions – 2016 ANES	13
IV Variable Distributions – 2015 YouGov	15
D Robustness Checks	16
I Controlling for Personality Characteristics	16
II Controlling for Individual Response Length	17
E Tables of Model Estimates	18
I Main Analyses	18
II Robustness Checks	25

*Ph.D. Candidate, Stony Brook University, patrick.kraft@stonybrook.edu.

Appendix A Detailed Information on Open-Ended Responses and Discursive Sophistication Components

I Distribution of Word Counts in Open-Ended Responses

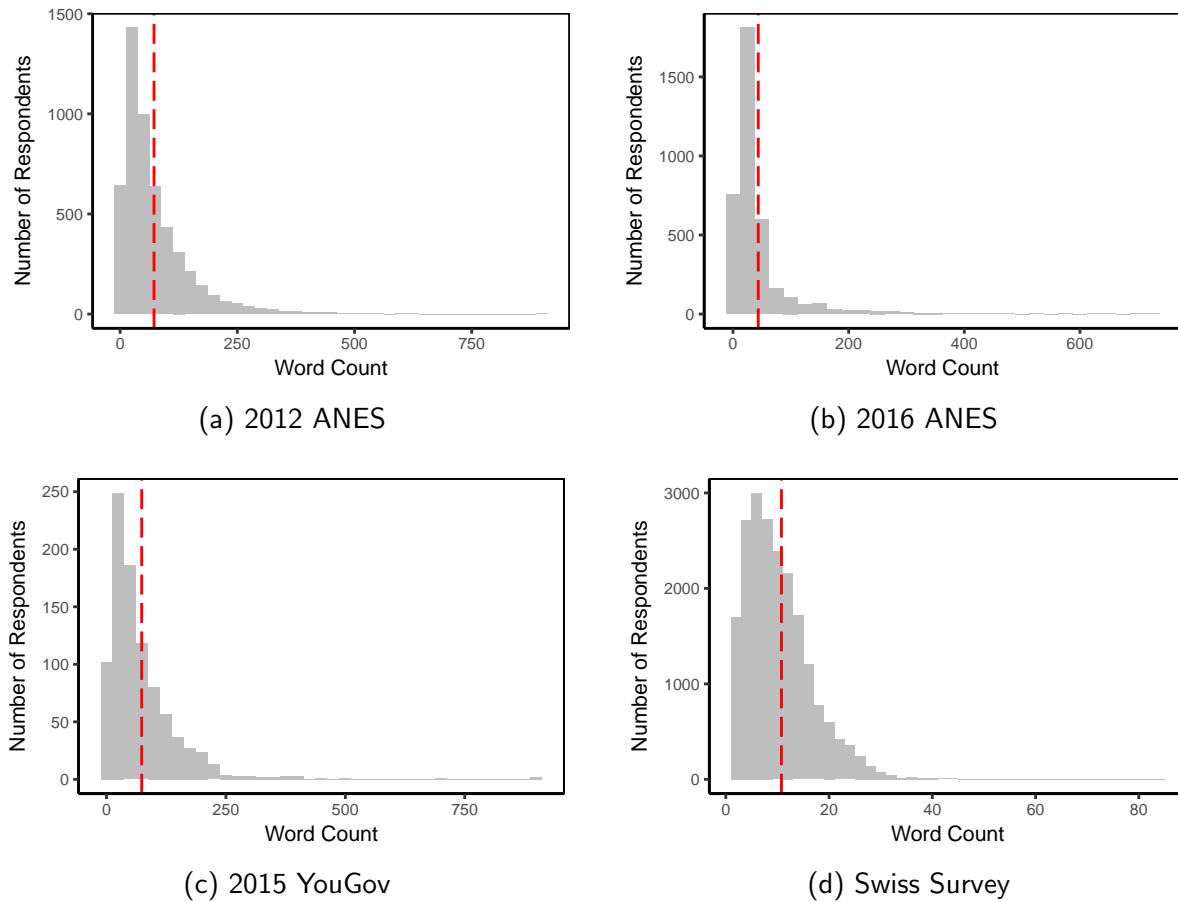


Figure A.1: Histograms of total word count in the collection of open-ended responses for each individual. The dashed red lines indicate the average response lengths in each survey.

II Overview of Topic Proportions

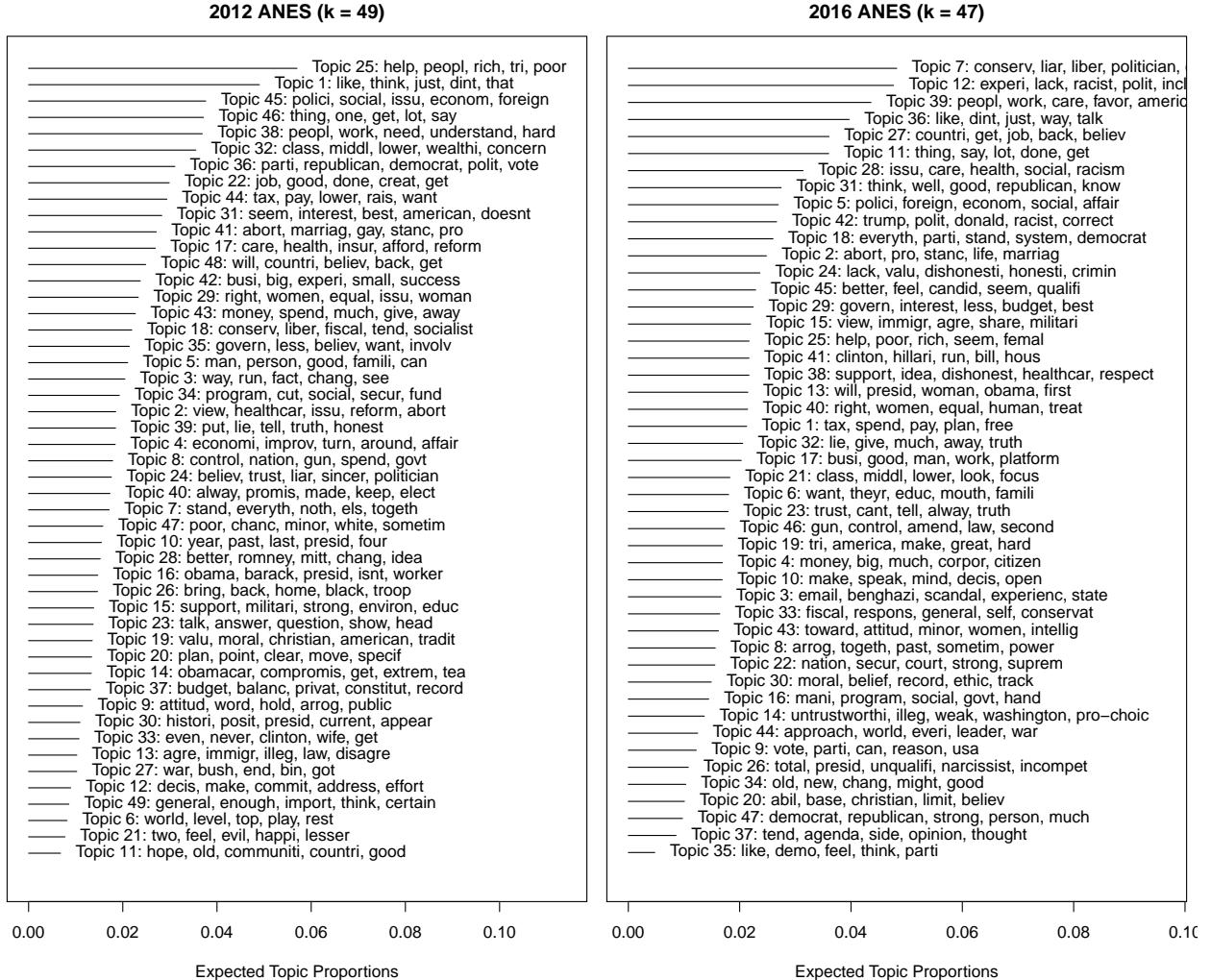


Figure A.2: Estimated topic proportions in the 2012 and 2016 ANES based on the structural topic model. See Appendix B for details on the model specification.

YouGov Survey (k = 47)

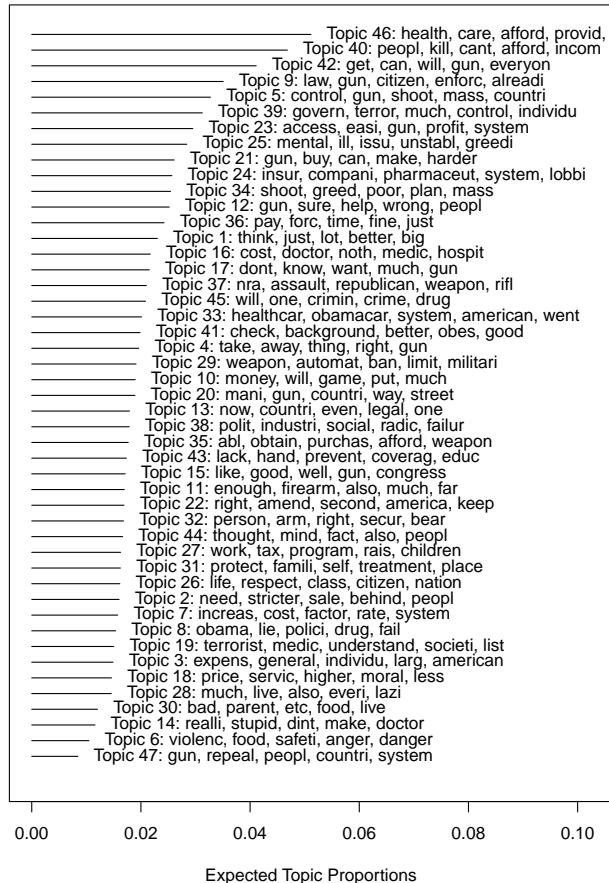


Figure A.3: Estimated topic proportions in the 2015 YouGov survey based on the structural topic model. See Appendix B for details on the model specification.

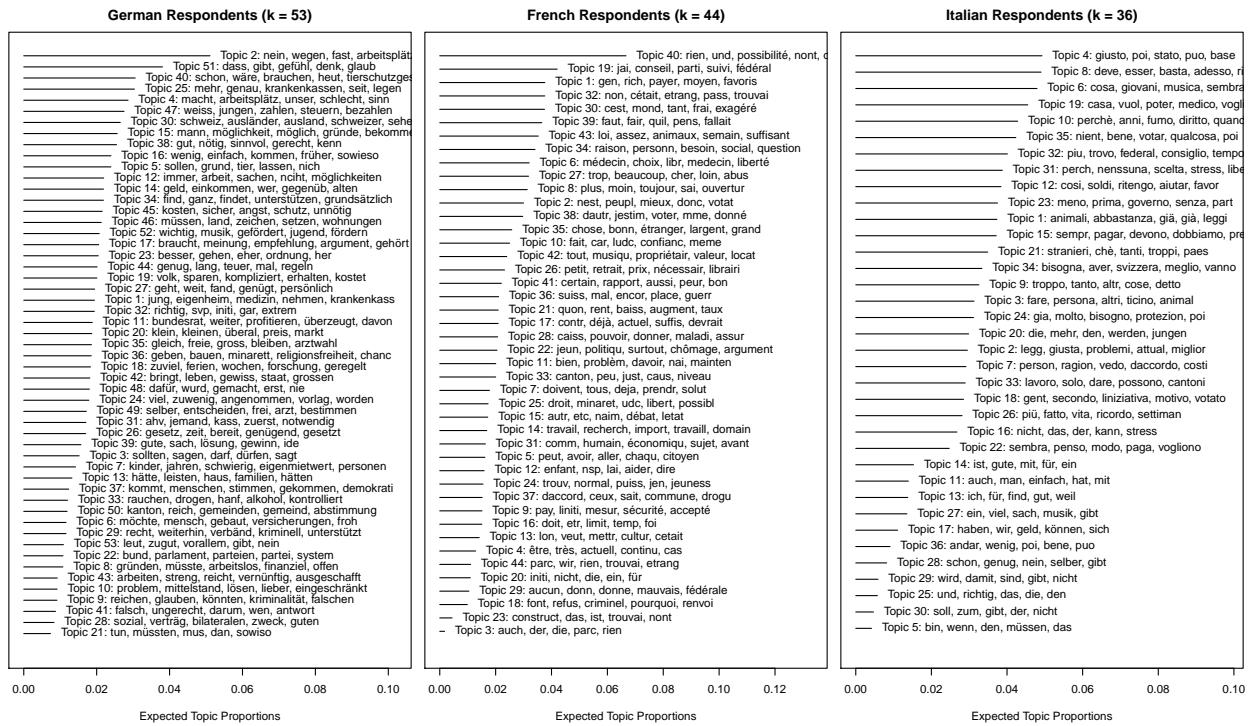


Figure A.4: Estimated topic proportions in the Swiss survey based on the structural topic model. See Appendix B for details on the model specification.

III Discursive Sophistication Components

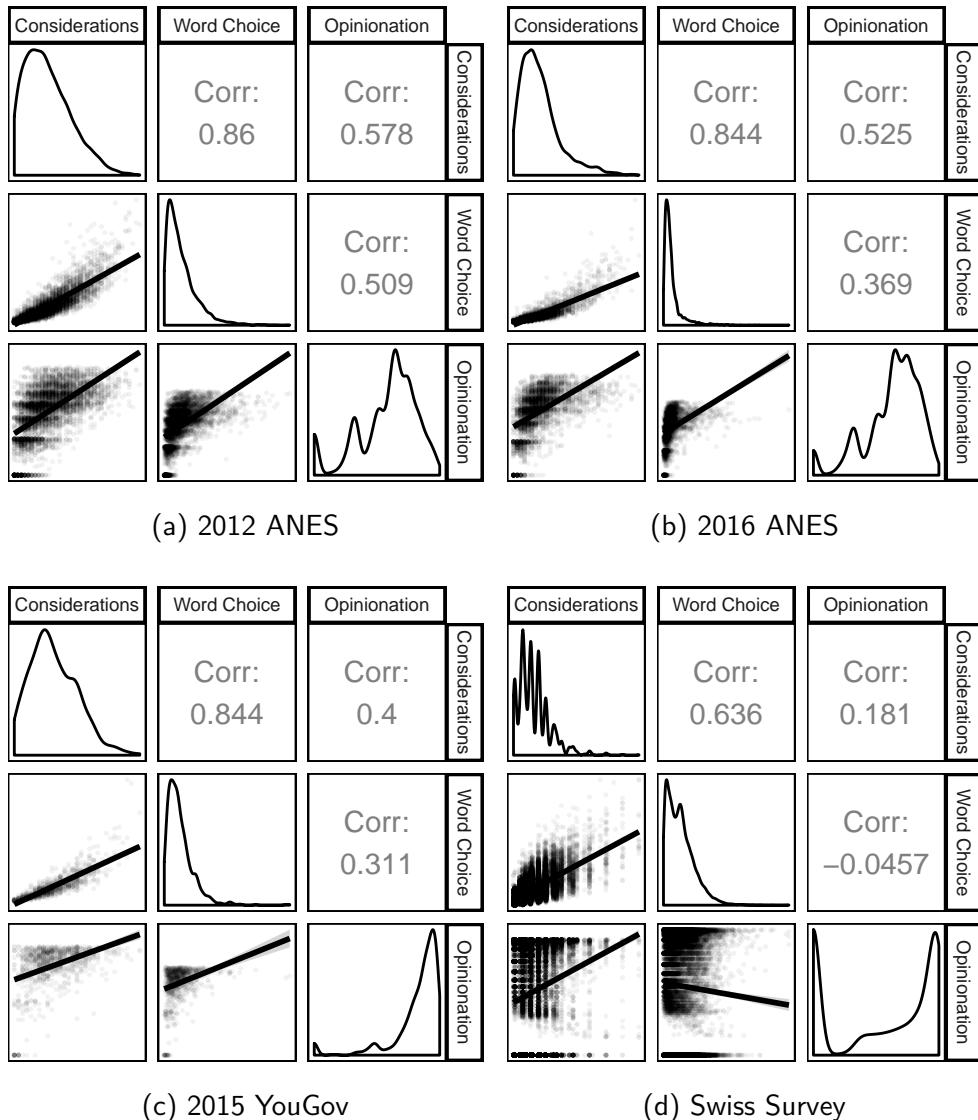


Figure A.5: Correlation matrix of individual components of discursive sophistication. The plots on the diagonal display univariate densities for each component. The panels in the lower triangular display the scatter plot of two measures as well as a linear fit.

Appendix B Pre-Processing and Topic Model Specification

I PreText Analysis

Two components of discursive sophistication (*considerations* and *word choice*) rely on quantities extracted from structural topic models (Roberts et al., 2014). As with any other text-as-data approach, a necessary first step before estimating the topic model is to pre-process the raw text and convert it into a document term matrix (DTM, see for example Manning et al., 2008). Common pre-processing procedures include stemming and lowercasing, as well as the removal of numbers, punctuation, stopwords, and infrequent terms. However, topic models and other unsupervised learning techniques can be sensitive to these pre-processing choices (c.f., Denny and Spirling, 2018). To address this issue, Denny and Spirling (2018) recommend that researchers compare DTMs under all possible pre-processing regimes. The authors propose *preText scores* as a measure to quantify the extent to which varying pre-processing regimes may yield unusual results compared to a baseline without any pre-processing.

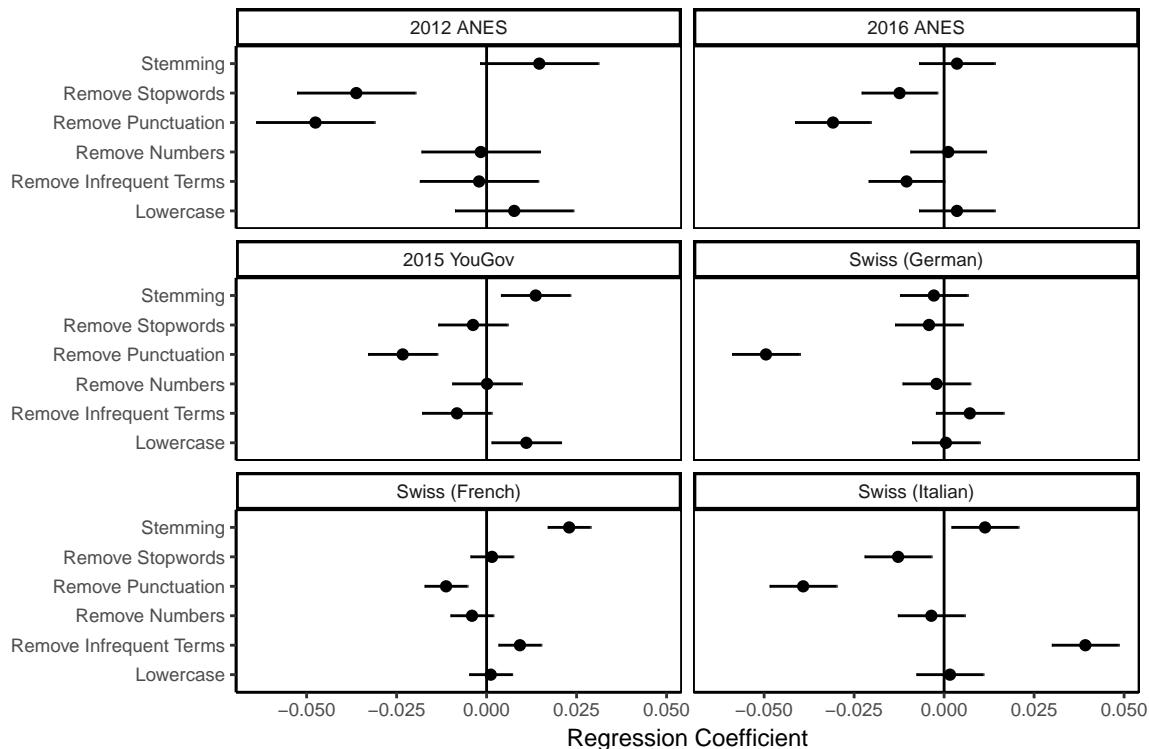


Figure B.1: PreText analysis of pre-processing decisions of open-ended responses across all datasets. Regression coefficients display the effects of each of the six pre-processing choices on the resulting preText score.

II Robustness Checks for Varying Model Specifications

Following the procedure outlined in Denny and Spirling (2018), Figure B.1 displays the results of a linear model regressing preText scores resulting from all possible pre-processing regimes on each individual step for a random subset of 500 open-ended responses in each of the studies included in the analyses. Significant coefficients indicate that the topic model results may be sensitive to the respective pre-processing step. As such, removing stopwords and punctuation, as well as removing infrequent terms (at least in the 2016 ANES) might be problematic. Denny and Spirling (2018), however, emphasize that the most important consideration in choosing pre-processing steps are theoretical. Given that the purpose of the topic model is to extract considerations related to political preferences, there are strong theoretical reasons to remove stopwords and punctuation from open-ended responses as they do not contain any relevant content. Furthermore, I apply lowercasing and stemming of terms to reduce resulting document term matrix to a computationally more manageable size and since these pre-processing steps should not be influential according to the preText analysis.

It is less obvious from a theoretical perspective whether to remove infrequent terms from open-ended responses, although it is preferred in order to make the estimation of the discursive sophistication components computationally efficient. Since the preText analysis for the 2016 ANES suggests that this pre-processing step might be influential, I compare discursive sophistication for both alternative regimes below (c.f., Denny and Spirling, 2018). Before turning to this sensitivity check, however, I consider another crucial modeling choice when working with topic models: determining the total number of topics k to be estimated. For all analyses reported below, the number of topics was selected using the algorithm proposed by Lee and Mimno (2014) and implemented in the `stm` package in R (Roberts, Stewart, and Tingley, 2014).¹

Figure B.2 examines whether the proposed measure of discursive sophistication is sensitive to the removal of infrequent terms as well as the chosen number of topics k . The y-axis depicts the preferred pre-processing regime including all steps discussed above while the x-axis plots results for alternative specifications. The panels on the left compare the preferred specification to discursive sophistication based on a reduced number of topics ($k = 20$). The middle panels additionally include infrequent terms instead of removing them.² The panels on the right do not perform stemming as part of the pre-processing step. Across all panels, discursive sophistication scores are highly correlated and therefore insensitive to pre-processing choices and varying numbers of topics.

In summary, open-ended responses in the analyses reported in the main text are pre-processed by stemming and lowercasing, as well as the removing numbers, punctuation, stopwords, and infrequent terms (i.e., terms that appear in fewer than 10 responses).³ While the results discussed in the manuscript are based on this preferred specification, the substantive results are robust for alternative pre-processing regimes or varying numbers of topics.

¹I used measures for age, gender, education, party identification, as well as an interaction between education and party identification as covariates for topic prevalence. This variable selection—with the exception of including gender—is equivalent to the procedure model specification described in Roberts et al. (2014).

²Calculating discursive sophistication with large numbers of topics while including infrequent terms is computationally prohibitive.

³Prior to applying these pre-processing steps, open-ended responses in the 2012 & 2016 ANES as well as the 2015 YouGov survey are cleaned by correcting spelling errors using an implementation of the Aspell spell-checking algorithm (www.aspell.net).

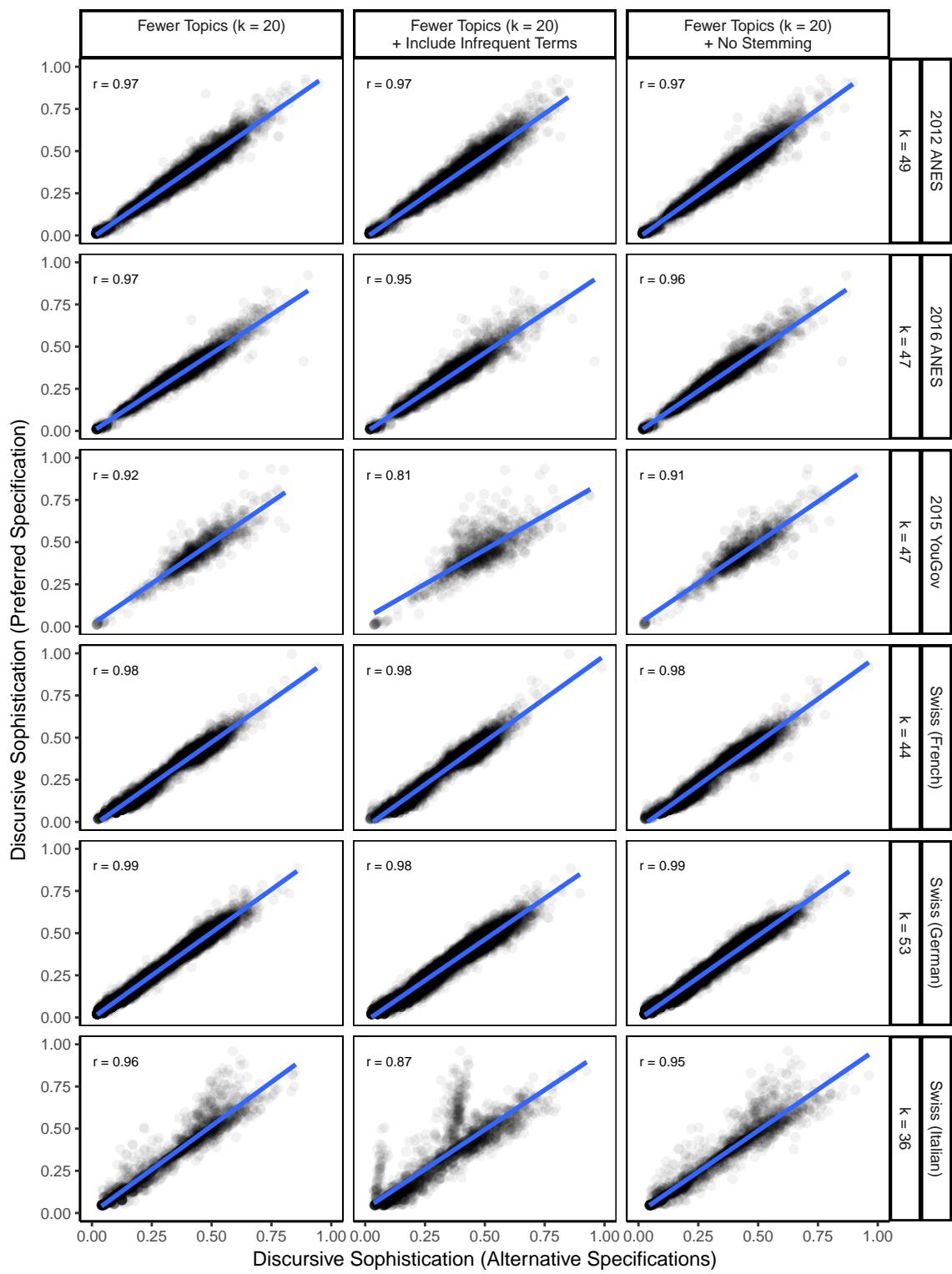


Figure B.2: Robustness of discursive sophistication measure for different pre-processing choices and topic model specifications.

Appendix C Additional Information on Remaining Variables

I Item Selection and Recoding

Conventional measures of political knowledge:

- *2012 ANES*: Additive index of correct responses to 5 knowledge items included in the pre-election wave (number of Presidential terms, size of budget deficit, length of Senate term, meaning of Medicare, federal government spending). ‘Don’t know’ responses are considered incorrect. Interviewer evaluations are based on the assessment of the respondent’s general level of information about politics recorded at the end of the pre-election wave.
- *2016 ANES*: Additive index of correct responses to 4 knowledge items included in the pre-election wave (length of Senate term, federal government spending, majority in House, majority in Senate). ‘Don’t know’ responses are considered incorrect. Interviewer evaluations are based on the assessment of the respondent’s general level of information about politics recorded at the end of the pre-election wave.
- *2015 YouGov Survey*: Additive index of correct responses to 8 knowledge items (Speaker of the House, meaning of TPP, Chair of Federal Reserve Board, current unemployment rate, Presidential veto override, meaning of Common Core, leading source of electricity in US, majority in Senate). ‘Don’t know’ responses are considered incorrect.

Dependent variables:

- *Turnout* (2012 & 2016 ANES): Dichotomous indicator, based on post-election wave.
- *Non-conventional participation* (2012 & 2016 ANES): Additive index of different forms of political engagement (participated in protest, signed petition, wore campaign button, wrote letter to public official).
- *Internal efficacy* (2012 & 2016 ANES): Sum of two agree/disagree items (politics too complicated, good understanding of political issues [reversed]).
- *External efficacy* (2012 & 2016 ANES): Sum of two agree/disagree items (public officials don’t care, people have no say about what the government does).
- *Information retrieval* (2015 YouGov Survey): Additive index of correct answers to 9 questions about the fictional disease described in the news article (symptoms: fatigue, headaches, diarrhea, joint pain, boils, warts, fever; virus spread; cure for the virus)
- *Candidate policy positions* (2012 & 2016 ANES): Placement of Republican and Democratic Presidential candidates on 7-point scale (ideology, government spending, defense spending, insurance policy, job guarantee, aid to Blacks, environment vs jobs).

Control variables:

- *Gender* (2012 & 2016 ANES, 2015 YouGov Survey): Dichotomous indicator for female respondents.
- *Wordsum vocabulary scores* (2012 & 2016 ANES): Modified version of the GSS wordsum vocabulary test consisting of 10 terms.
- *Media exposure* (2012 & 2016 ANES): Additive index of the frequency of weekly exposure to various political information sources such as newspapers or TV news (2012 ANES). In the 2016 ANES, it only consists of a single item measuring the number of days in the past week the respondent has spent watching/reading/listening news on any media.
- *Political discussion frequency* (2012 & 2016 ANES): Self-reported count of days in the past week spent discussing politics with family or friends.
- *College education* (2012 & 2016 ANES, 2015 YouGov Survey): Dichotomous indicator for Bachelor's degree or higher.
- *Family/Household income* (2012 & 2016 ANES, 2015 YouGov Survey): Self-reported household income categories.
- *Age* (2012 & 2016 ANES, 2015 YouGov Survey): Logged age in years.
- *Race* (2012 & 2016 ANES, 2015 YouGov Survey): Dichotomous indicator for black non-Hispanic vs. others.
- *Church attendance* (2012 & 2016 ANES, 2015 YouGov Survey): Six-category indicator of the frequency of church attendance.
- *Survey Mode* (2012 & 2016 ANES): Dichotomous indicator for face-to-face vs. online samples of the ANES surveys.
- *Personality characteristics* (2012 & 2016 ANES): Measures of extraversion and being reserved, part of the Ten Item Personality Inventory (TIPI) measuring the “Big Five” personality traits.
- *Response length* (2012 & 2016 ANES): Logged number of words in the collection of open-ended responses by each individual.

II Variable Distributions – 2012 ANES

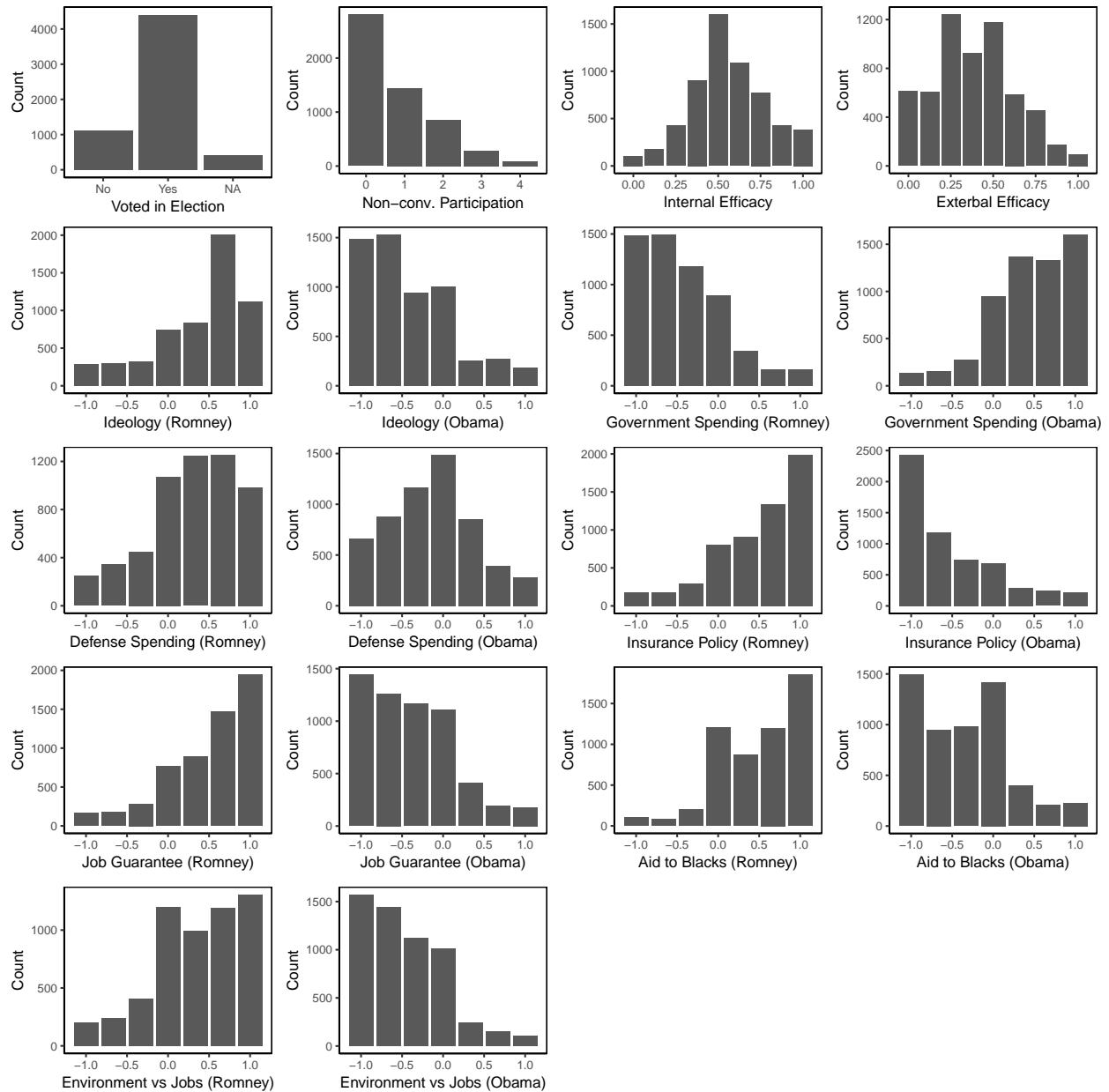


Figure C.1: Histograms of dependent variables included in 2012 ANES.

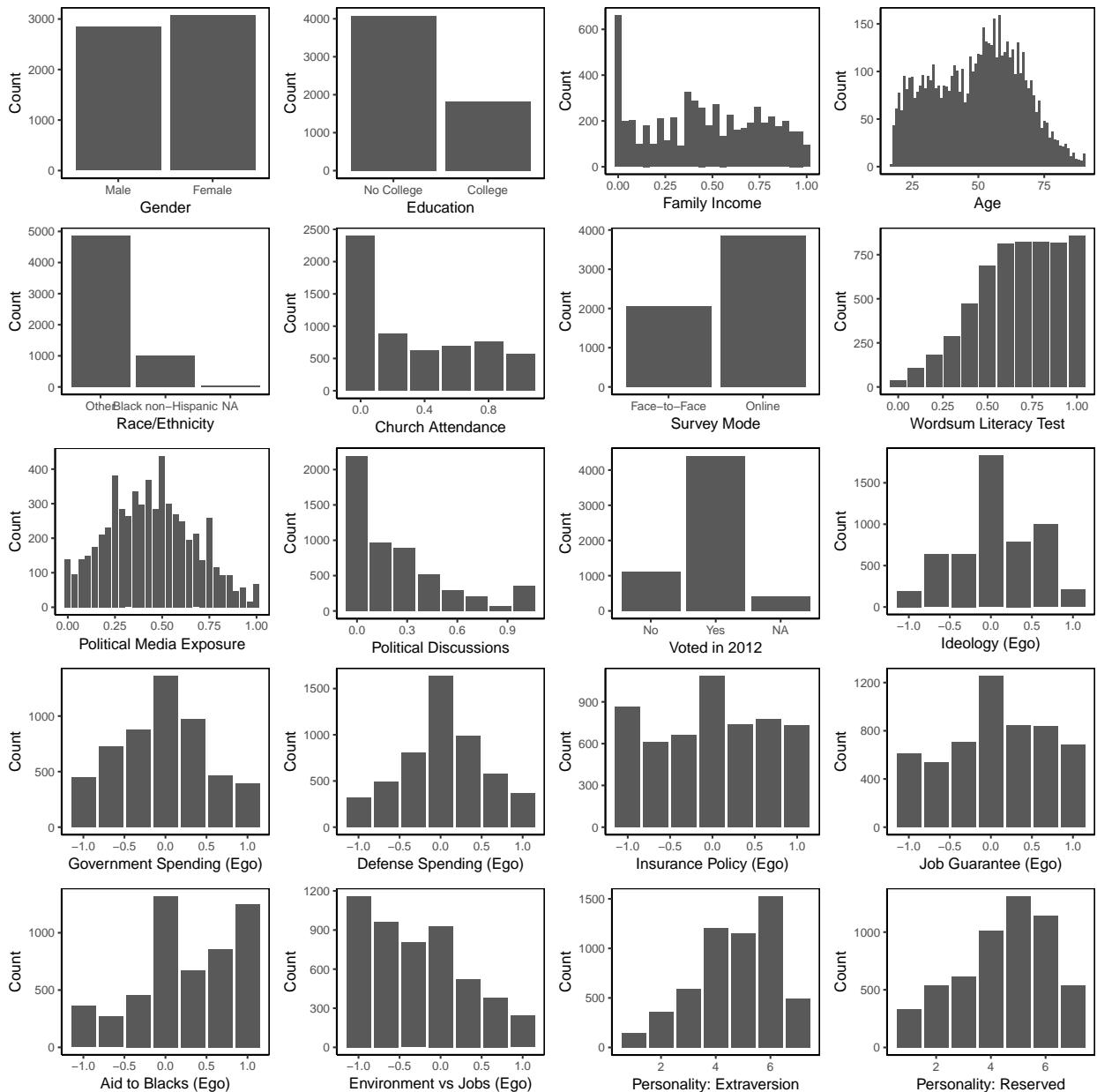


Figure C.2: Histograms of independent variables included in 2012 ANES.

III Variable Distributions – 2016 ANES

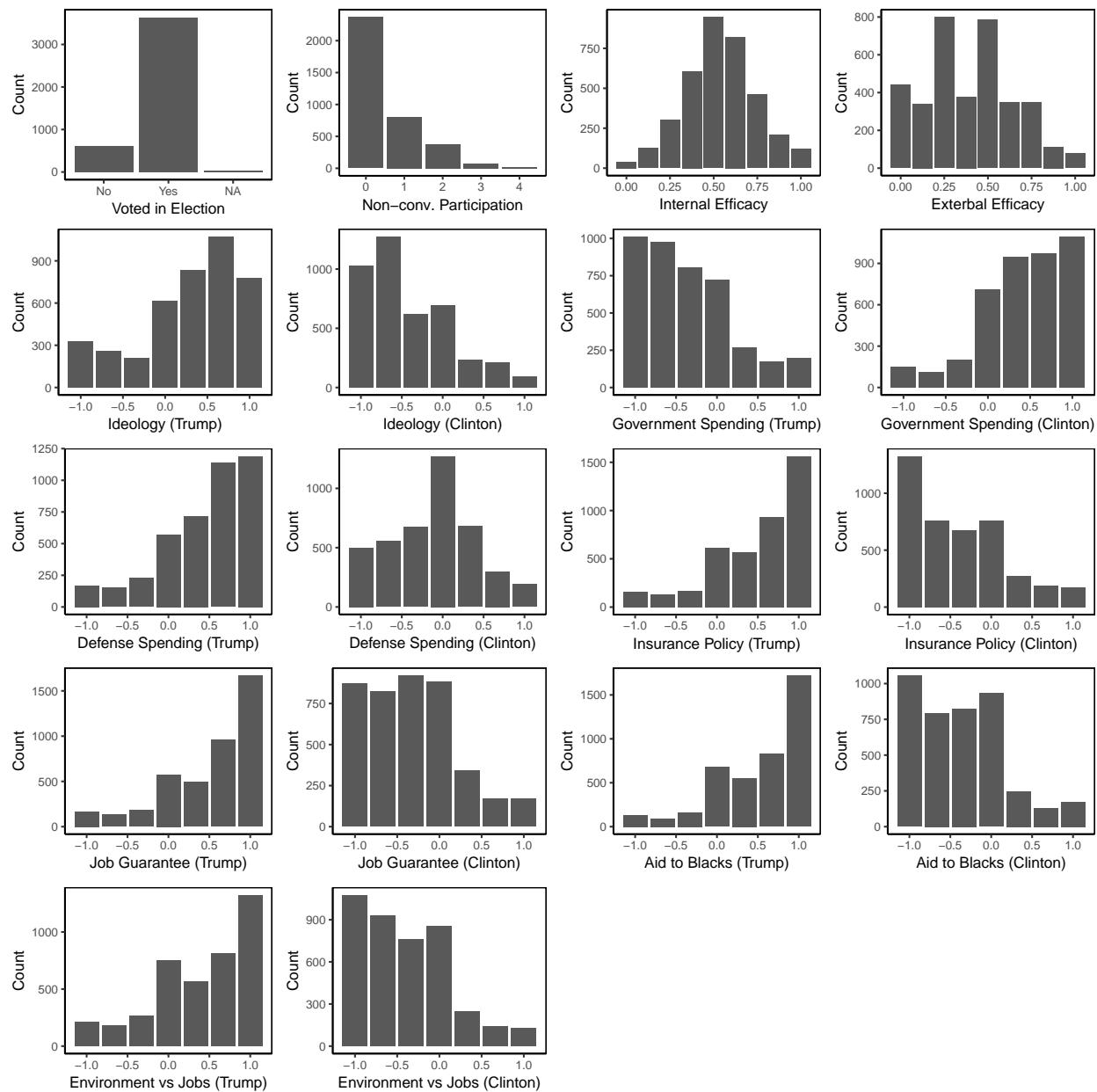


Figure C.3: Histograms of dependent variables included in 2016 ANES.

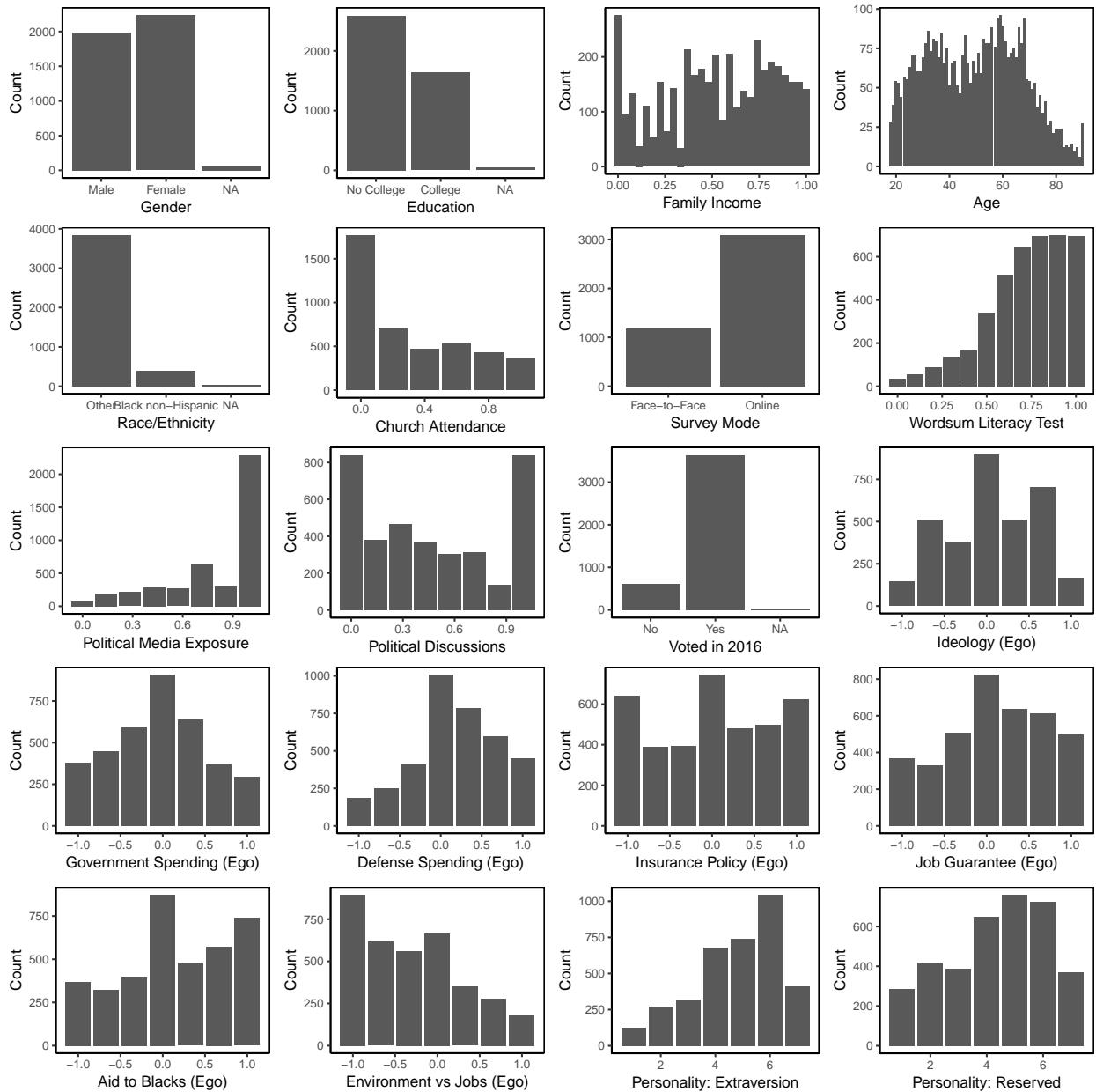


Figure C.4: Histograms of independent variables included in 2016 ANES.

IV Variable Distributions – 2015 YouGov

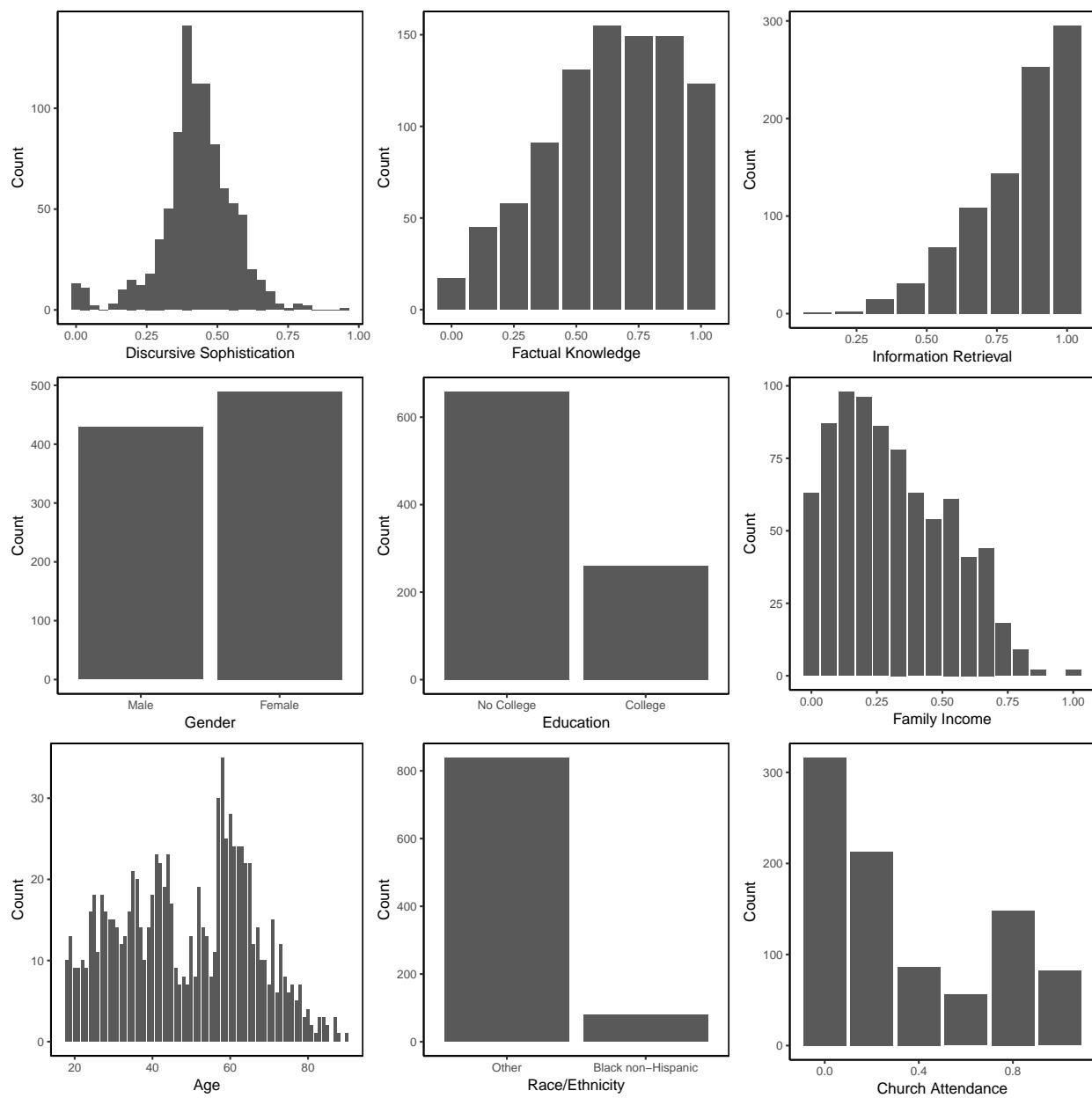


Figure C.5: Histogram of variables included in 2015 YouGov survey.

Appendix D Robustness Checks

I Controlling for Personality Characteristics

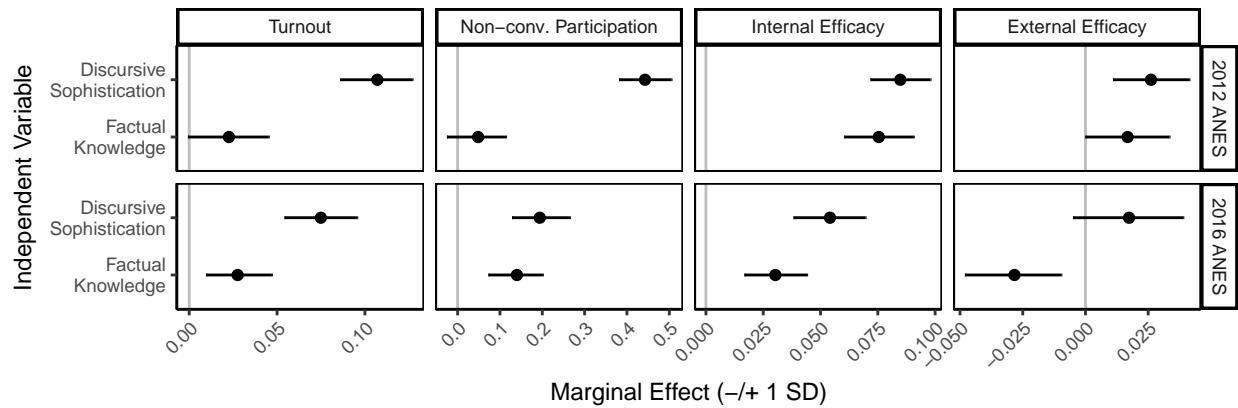


Figure D.1: Effects of sophistication on internal efficacy, external efficacy, non-conventional participation, and turnout in the 2012 and 2016 ANES. For each dependent variable, the figure displays the change in expected values after increasing each sophistication measure from -1 to +1 standard deviation from its mean (including 95% confidence intervals). Model estimates are based on logistic regression (turnout) or OLS (internal efficacy, external efficacy, non-conventional participation). Both sophistication measure are included simultaneously while controlling for gender, education, income, age, race, church attendance, survey mode, Wordsum vocabulary scores, as well as personality characteristics (extraversion and being reserved).

II Controlling for Individual Response Length

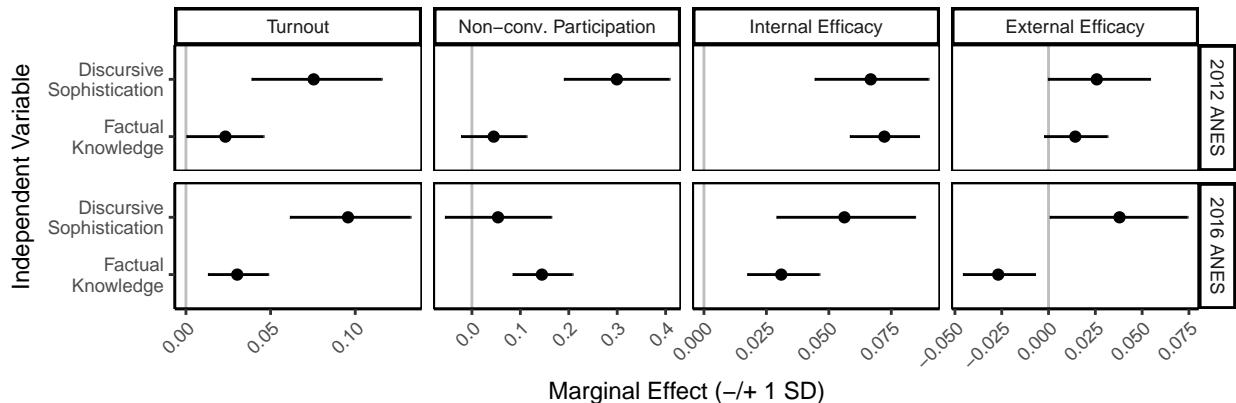


Figure D.2: Effects of sophistication on internal efficacy, external efficacy, non-conventional participation, and turnout in the 2012 and 2016 ANES. For each dependent variable, the figure displays the change in expected values after increasing each sophistication measure from -1 to $+1$ standard deviation from its mean (including 95% confidence intervals). Model estimates are based on logistic regression (turnout) or OLS (internal efficacy, external efficacy, non-conventional participation). Both sophistication measure are included simultaneously while controlling for gender, education, income, age, race, church attendance, survey mode, Wordsum vocabulary scores, as well as the logged word count across open-ended responses.

Appendix E Tables of Model Estimates

I Main Analyses

Table E.1: Effects of sophistication on turnout, non-conventional participation, internal efficacy, and external efficacy in the 2012 ANES. Standard errors in parentheses. Estimates are used for Figure 2 in the main text.

	<i>Dependent variable:</i>			
	Turnout	Participation	Internal Efficacy	External Efficacy
Discursive Soph.	2.921*** (0.299)	1.440*** (0.098)	0.278*** (0.020)	0.084*** (0.024)
Factual Knowledge	0.432* (0.218)	0.099 (0.075)	0.158*** (0.016)	0.032 (0.018)
Female	0.086 (0.085)	-0.067* (0.028)	-0.053*** (0.006)	0.016* (0.007)
College Degree	0.350** (0.112)	0.159*** (0.034)	0.022** (0.007)	0.035*** (0.008)
Family Income	0.947*** (0.156)	0.022 (0.052)	0.010 (0.011)	0.016 (0.013)
Age (log)	0.988*** (0.105)	0.102** (0.038)	-0.006 (0.008)	-0.014 (0.009)
African American	0.910*** (0.123)	0.096* (0.038)	0.066*** (0.008)	0.082*** (0.009)
Church Attendance	0.752*** (0.129)	0.112** (0.040)	0.010 (0.008)	0.048*** (0.010)
Mode: Online	0.530*** (0.094)	0.227*** (0.033)	0.017* (0.007)	-0.039*** (0.008)
Wordsum Score	0.638** (0.219)	0.403*** (0.076)	0.092*** (0.016)	0.013 (0.019)
Constant	-5.019*** (0.401)	-0.598*** (0.145)	0.326*** (0.030)	0.352*** (0.035)
Observations	4,716	4,692	4,996	4,985
R ²		0.124	0.161	0.043
Log Likelihood	-1,868.199			

Note:

*p<0.05; **p<0.01; ***p<0.001

Table E.2: Effects of sophistication on turnout, non-conventional participation, internal efficacy, and external efficacy in the 2016 ANES. Standard errors in parentheses. Estimates are used for Figure 2 in the main text.

	<i>Dependent variable:</i>			
	Turnout	Participation	Internal Efficacy	External Efficacy
Discursive Soph.	3.891*** (0.487)	0.792*** (0.131)	0.229*** (0.031)	0.076 (0.041)
Factual Knowledge	0.730*** (0.219)	0.269*** (0.058)	0.058*** (0.014)	-0.051** (0.018)
Female	0.177 (0.114)	0.061* (0.029)	-0.059*** (0.007)	-0.003 (0.009)
College Degree	0.554*** (0.141)	0.092** (0.033)	0.058*** (0.008)	0.057*** (0.010)
Family Income	0.487* (0.207)	-0.077 (0.055)	0.020 (0.013)	0.062*** (0.017)
Age (log)	0.840*** (0.139)	-0.110** (0.038)	0.020* (0.009)	-0.005 (0.012)
African American	1.126*** (0.237)	0.097 (0.051)	0.057*** (0.012)	-0.020 (0.016)
Church Attendance	1.067*** (0.191)	-0.175*** (0.043)	-0.006 (0.010)	0.080*** (0.014)
Mode: Online	0.192 (0.136)	0.110** (0.036)	0.069*** (0.009)	-0.016 (0.011)
Wordsum Score	0.931*** (0.275)	0.402*** (0.079)	0.103*** (0.019)	0.017 (0.025)
Constant	-4.276*** (0.535)	0.213 (0.151)	0.231*** (0.036)	0.338*** (0.048)
Observations	3,486	3,039	3,037	3,038
R ²		0.062	0.147	0.043
Log Likelihood	-1,085.646			

Note:

*p<0.05; **p<0.01; ***p<0.001

Table E.3: Effects of sophistication on information retrieval in the 2015 YouGov study. Standard errors in parentheses. Estimates are used for Figure 3 in the main text.

	<i>Dependent variable:</i> Information Retrieval
Discursive Soph.	0.247*** (0.045)
Factual Knowledge	0.133*** (0.027)
Female	0.044*** (0.012)
College Degree	0.015 (0.014)
Family Income	−0.062* (0.030)
Age (log)	0.032* (0.016)
African American	−0.026 (0.021)
Church Attendance	−0.059*** (0.017)
Constant	0.534*** (0.061)
Observations	792
R ²	0.136

Note: *p<0.05; **p<0.01; ***p<0.001

Table E.4: Error variance reduction in candidate placements on multiple issues in the 2012 ANES. Estimates are used for Figure 4 in the main text.

Policy Position	Candidate	Independent Var.	$E[\gamma]$	$sd(\gamma)$	95% Cred. Int.	\hat{R}
Ideology	Romney	Discursive Soph.	-0.71	0.07	(-0.856; -0.562)	1.00
Ideology	Romney	Factual Knowl.	-0.65	0.06	(-0.755; -0.538)	1.00
Ideology	Obama	Discursive Soph.	-0.41	0.07	(-0.544; -0.264)	1.00
Ideology	Obama	Factual Knowl.	-0.58	0.06	(-0.69; -0.47)	1.00
Government Spending	Romney	Discursive Soph.	-0.34	0.07	(-0.482; -0.205)	1.00
Government Spending	Romney	Factual Knowl.	-0.41	0.06	(-0.52; -0.303)	1.00
Government Spending	Obama	Discursive Soph.	-0.54	0.07	(-0.68; -0.398)	1.00
Government Spending	Obama	Factual Knowl.	-0.47	0.06	(-0.582; -0.354)	1.00
Defense Spending	Romney	Discursive Soph.	-0.18	0.08	(-0.326; -0.031)	1.00
Defense Spending	Romney	Factual Knowl.	-0.33	0.06	(-0.443; -0.223)	1.00
Defense Spending	Obama	Discursive Soph.	-0.17	0.08	(-0.318; -0.028)	1.00
Defense Spending	Obama	Factual Knowl.	-0.19	0.05	(-0.287; -0.086)	1.00
Insurance Policy	Romney	Discursive Soph.	-0.26	0.07	(-0.398; -0.12)	1.00
Insurance Policy	Romney	Factual Knowl.	-0.22	0.05	(-0.326; -0.116)	1.00
Insurance Policy	Obama	Discursive Soph.	-0.33	0.07	(-0.476; -0.194)	1.00
Insurance Policy	Obama	Factual Knowl.	-0.41	0.06	(-0.515; -0.297)	1.00
Job Guarantee	Romney	Discursive Soph.	-0.59	0.08	(-0.734; -0.436)	1.00
Job Guarantee	Romney	Factual Knowl.	-0.33	0.06	(-0.439; -0.223)	1.00
Job Guarantee	Obama	Discursive Soph.	-0.35	0.07	(-0.498; -0.215)	1.00
Job Guarantee	Obama	Factual Knowl.	-0.20	0.05	(-0.304; -0.091)	1.00
Aid to Blacks	Romney	Discursive Soph.	-0.22	0.07	(-0.368; -0.069)	1.00
Aid to Blacks	Romney	Factual Knowl.	-0.27	0.05	(-0.369; -0.163)	1.00
Aid to Blacks	Obama	Discursive Soph.	-0.21	0.07	(-0.36; -0.07)	1.00
Aid to Blacks	Obama	Factual Knowl.	-0.20	0.06	(-0.312; -0.095)	1.00
Environment vs Jobs	Romney	Discursive Soph.	-0.29	0.08	(-0.438; -0.138)	1.00
Environment vs Jobs	Romney	Factual Knowl.	-0.34	0.06	(-0.456; -0.227)	1.00
Environment vs Jobs	Obama	Discursive Soph.	-0.34	0.08	(-0.484; -0.181)	1.00
Environment vs Jobs	Obama	Factual Knowl.	-0.35	0.06	(-0.465; -0.233)	1.00

Table E.5: Error variance reduction in candidate placements on multiple issues in the 2016 ANES. Estimates are used for Figure 4 in the main text.

Policy Position	Candidate	Independent Var.	$E[\gamma]$	$sd(\gamma)$	95% Cred. Int.	\hat{R}
Ideology	Trump	Discursive Soph.	-0.77	0.11	(-0.984; -0.55)	1.00
Ideology	Trump	Factual Knowl.	-0.16	0.06	(-0.272; -0.057)	1.00
Ideology	Clinton	Discursive Soph.	-0.52	0.10	(-0.71; -0.313)	1.00
Ideology	Clinton	Factual Knowl.	-0.15	0.05	(-0.255; -0.048)	1.00
Government Spending	Trump	Discursive Soph.	-0.62	0.10	(-0.835; -0.424)	1.00
Government Spending	Trump	Factual Knowl.	-0.09	0.05	(-0.19; 0.013)	1.00
Government Spending	Clinton	Discursive Soph.	-1.00	0.11	(-1.213; -0.785)	1.00
Government Spending	Clinton	Factual Knowl.	-0.18	0.05	(-0.284; -0.079)	1.00
Defense Spending	Trump	Discursive Soph.	-0.37	0.10	(-0.563; -0.171)	1.00
Defense Spending	Trump	Factual Knowl.	-0.08	0.05	(-0.183; 0.021)	1.00
Defense Spending	Clinton	Discursive Soph.	-0.54	0.11	(-0.744; -0.319)	1.00
Defense Spending	Clinton	Factual Knowl.	-0.04	0.05	(-0.141; 0.061)	1.00
Insurance Policy	Trump	Discursive Soph.	-0.79	0.10	(-0.991; -0.579)	1.00
Insurance Policy	Trump	Factual Knowl.	-0.06	0.05	(-0.168; 0.04)	1.00
Insurance Policy	Clinton	Discursive Soph.	-0.74	0.11	(-0.959; -0.521)	1.00
Insurance Policy	Clinton	Factual Knowl.	-0.14	0.05	(-0.241; -0.033)	1.00
Job Guarantee	Trump	Discursive Soph.	-0.86	0.10	(-1.061; -0.671)	1.00
Job Guarantee	Trump	Factual Knowl.	-0.06	0.05	(-0.157; 0.043)	1.00
Job Guarantee	Clinton	Discursive Soph.	-0.76	0.10	(-0.965; -0.546)	1.00
Job Guarantee	Clinton	Factual Knowl.	-0.04	0.05	(-0.134; 0.059)	1.00
Aid to Blacks	Trump	Discursive Soph.	-0.66	0.10	(-0.862; -0.466)	1.00
Aid to Blacks	Trump	Factual Knowl.	-0.00	0.05	(-0.099; 0.097)	1.00
Aid to Blacks	Clinton	Discursive Soph.	-0.47	0.10	(-0.663; -0.277)	1.00
Aid to Blacks	Clinton	Factual Knowl.	-0.06	0.05	(-0.158; 0.045)	1.00
Environment vs Jobs	Trump	Discursive Soph.	-0.65	0.11	(-0.861; -0.443)	1.00
Environment vs Jobs	Trump	Factual Knowl.	-0.18	0.05	(-0.281; -0.072)	1.00
Environment vs Jobs	Clinton	Discursive Soph.	-0.76	0.10	(-0.956; -0.554)	1.00
Environment vs Jobs	Clinton	Factual Knowl.	-0.24	0.05	(-0.348; -0.142)	1.00

Table E.6: Effects of gender on discursive sophistication and factual knowledge in the 2012 ANES and 2016 ANES. Standard errors in parentheses. Estimates are used for Figure 7 in the main text.

	<i>Dependent variable:</i>			
	Discursive Sophistication		Factual Knowledge	
	2012 ANES	2016 ANES	2012 ANES	2016 ANES
Female	0.001 (0.004)	-0.002 (0.004)	-0.050*** (0.006)	-0.029** (0.009)
Media Exposure	0.042*** (0.010)	0.034*** (0.008)	0.067*** (0.014)	0.056** (0.019)
Discussion Frequency	0.121*** (0.008)	0.058*** (0.006)	0.062*** (0.010)	0.078*** (0.013)
College Degree	0.070*** (0.005)	0.045*** (0.004)	0.091*** (0.007)	0.045*** (0.010)
Family Income	0.054*** (0.008)	0.043*** (0.007)	0.116*** (0.010)	0.089*** (0.017)
Age (log)	0.017** (0.006)	0.003 (0.005)	0.081*** (0.008)	0.092*** (0.013)
African American	-0.010 (0.006)	-0.026*** (0.007)	-0.085*** (0.008)	-0.004 (0.016)
Church Attendance	0.004 (0.006)	0.005 (0.006)	-0.003 (0.008)	-0.006 (0.013)
Mode: Online	-0.044*** (0.005)	-0.110*** (0.004)	0.092*** (0.006)	0.085*** (0.010)
Constant	0.199*** (0.022)	0.284*** (0.020)	0.195*** (0.030)	0.009 (0.046)
Observations	4,690	3,119	4,690	3,108
R ²	0.178	0.265	0.282	0.110

Note:

*p<0.05; **p<0.01; ***p<0.001

Table E.7: Effects of gender on discursive sophistication and factual knowledge in the 2015 YouGov Study. Standard errors in parentheses. Estimates are used for Figure 7 in the main text.

	<i>Dependent variable:</i>	
	Discursive Sophistication	Factual Knowledge
Female	0.008 (0.009)	-0.157*** (0.016)
College Degree	0.027* (0.011)	0.108*** (0.019)
Family Income	0.003 (0.024)	0.233*** (0.041)
Age (log)	0.040** (0.013)	0.135*** (0.021)
African American	-0.064*** (0.017)	-0.086** (0.029)
Church Attendance	-0.008 (0.013)	-0.051* (0.023)
Constant	0.272*** (0.049)	0.125 (0.083)
Observations	792	792
R ²	0.043	0.272

Note:

*p<0.05; **p<0.01; ***p<0.001

II Robustness Checks

Table E.8: Effects of sophistication on turnout, non-conventional participation, internal efficacy, and external efficacy controlling for personality characteristics in the 2012 ANES. Standard errors in parentheses. Estimates are used for Figure D.1 in the appendix.

	<i>Dependent variable:</i>			
	Turnout	Participation	Internal Efficacy	External Efficacy
Discursive Soph.	2.912*** (0.300)	1.419*** (0.098)	0.272*** (0.021)	0.085*** (0.025)
Factual Knowledge	0.434* (0.218)	0.105 (0.075)	0.165*** (0.016)	0.036 (0.019)
Female	0.088 (0.085)	-0.078** (0.028)	-0.059*** (0.006)	0.009 (0.007)
College Degree	0.352** (0.112)	0.160*** (0.034)	0.020** (0.007)	0.033*** (0.008)
Family Income	0.941*** (0.157)	-0.002 (0.052)	0.006 (0.011)	0.010 (0.013)
Age (log)	0.981*** (0.106)	0.102** (0.039)	-0.003 (0.008)	-0.017 (0.010)
African American	0.902*** (0.124)	0.101** (0.038)	0.067*** (0.008)	0.078*** (0.010)
Church Attendance	0.732*** (0.130)	0.104** (0.040)	0.006 (0.008)	0.044*** (0.010)
Mode: Online	0.541*** (0.096)	0.249*** (0.033)	0.022** (0.007)	-0.033*** (0.008)
Wordsum Score	0.649** (0.220)	0.409*** (0.076)	0.092*** (0.016)	0.021 (0.019)
Extraversion	0.029 (0.030)	0.032** (0.010)	0.009*** (0.002)	0.020*** (0.002)
Reserved	0.009 (0.026)	-0.023** (0.009)	-0.005** (0.002)	0.002 (0.002)
Constant	-5.181*** (0.449)	-0.645*** (0.158)	0.296*** (0.034)	0.255*** (0.039)
Observations	4,691	4,685	4,691	4,679
R ²		0.129	0.171	0.058
Log Likelihood	-1,861.790			

Note:

*p<0.05; **p<0.01; ***p<0.001

Table E.9: Effects of sophistication on turnout, non-conventional participation, internal efficacy, and external efficacy controlling for personality characteristics in the 2016 ANES. Standard errors in parentheses. Estimates are used for Figure D.1 in the appendix.

	<i>Dependent variable:</i>			
	Turnout	Participation	Internal Efficacy	External Efficacy
Discursive Soph.	3.936*** (0.543)	0.760*** (0.132)	0.213*** (0.031)	0.070 (0.042)
Factual Knowledge	0.717** (0.241)	0.260*** (0.059)	0.056*** (0.014)	-0.052** (0.019)
Female	0.172 (0.125)	0.046 (0.029)	-0.064*** (0.007)	-0.004 (0.009)
College Degree	0.616*** (0.155)	0.085** (0.033)	0.056*** (0.008)	0.055*** (0.010)
Family Income	0.471* (0.230)	-0.085 (0.055)	0.017 (0.013)	0.060*** (0.017)
Age (log)	0.894*** (0.153)	-0.119** (0.039)	0.018* (0.009)	-0.010 (0.012)
African American	1.228*** (0.267)	0.102* (0.052)	0.057*** (0.012)	-0.026 (0.016)
Church Attendance	0.930*** (0.205)	-0.172*** (0.043)	-0.007 (0.010)	0.078*** (0.014)
Mode: Online	0.199 (0.148)	0.111** (0.036)	0.069*** (0.009)	-0.015 (0.011)
Wordsum Score	0.999** (0.311)	0.436*** (0.080)	0.114*** (0.019)	0.022 (0.025)
Extraversion	0.090* (0.041)	0.023* (0.010)	0.011*** (0.002)	0.012*** (0.003)
Reserved	0.005 (0.037)	-0.013 (0.009)	-0.002 (0.002)	0.005 (0.003)
Constant	-4.948*** (0.646)	0.196 (0.163)	0.199*** (0.039)	0.277*** (0.051)
Observations	2,999	3,011	3,012	3,012
R ²		0.065	0.155	0.047
Log Likelihood	-907.430			

Note:

*p<0.05; **p<0.01; ***p<0.001

Table E.10: Effects of sophistication on turnout, non-conventional participation, internal efficacy, and external efficacy controlling for individual response length in the 2012 ANES. Standard errors in parentheses. Estimates are used for Figure D.2 in the appendix.

	<i>Dependent variable:</i>			
	Turnout	Participation	Internal Efficacy	External Efficacy
Discursive Soph.	2.053*** (0.543)	0.967*** (0.186)	0.215*** (0.039)	0.082 (0.045)
Factual Knowledge	0.426 (0.218)	0.096 (0.075)	0.158*** (0.016)	0.032 (0.018)
Female	0.067 (0.085)	-0.076** (0.028)	-0.054*** (0.006)	0.016* (0.007)
College Degree	0.370** (0.113)	0.167*** (0.034)	0.023** (0.007)	0.035*** (0.008)
Family Income	0.948*** (0.156)	0.026 (0.052)	0.011 (0.011)	0.016 (0.013)
Age (log)	0.976*** (0.106)	0.094* (0.039)	-0.007 (0.008)	-0.014 (0.009)
African American	0.896*** (0.123)	0.088* (0.038)	0.064*** (0.008)	0.082*** (0.009)
Church Attendance	0.758*** (0.129)	0.113** (0.040)	0.010 (0.008)	0.048*** (0.010)
Mode: Online	0.574*** (0.097)	0.246*** (0.034)	0.020** (0.007)	-0.039*** (0.008)
Wordsum Score	0.626** (0.219)	0.396*** (0.076)	0.091*** (0.016)	0.013 (0.019)
Word Count (log)	0.970 (0.507)	0.539** (0.180)	0.071 (0.037)	0.003 (0.044)
Constant	-5.237*** (0.418)	-0.710*** (0.150)	0.311*** (0.031)	0.352*** (0.036)
Observations	4,716	4,692	4,996	4,985
R ²		0.126	0.162	0.043
Log Likelihood	-1,866.366			

Note:

*p<0.05; **p<0.01; ***p<0.001

Table E.11: Effects of sophistication on turnout, non-conventional participation, internal efficacy, and external efficacy controlling for individual response length in the 2016 ANES. Standard errors in parentheses. Estimates are used for Figure D.2 in the appendix.

	<i>Dependent variable:</i>			
	Turnout	Participation	Internal Efficacy	External Efficacy
Discursive Soph.	4.762*** (0.843)	0.224 (0.231)	0.222*** (0.055)	0.153* (0.073)
Factual Knowledge	0.727*** (0.219)	0.269*** (0.058)	0.058*** (0.014)	-0.051** (0.018)
Female	0.189 (0.114)	0.054 (0.029)	-0.059*** (0.007)	-0.002 (0.009)
College Degree	0.547*** (0.141)	0.096** (0.033)	0.058*** (0.008)	0.056*** (0.010)
Family Income	0.487* (0.207)	-0.074 (0.055)	0.020 (0.013)	0.062*** (0.017)
Age (log)	0.842*** (0.139)	-0.111** (0.038)	0.020* (0.009)	-0.005 (0.012)
African American	1.137*** (0.237)	0.090 (0.051)	0.057*** (0.012)	-0.019 (0.016)
Church Attendance	1.057*** (0.191)	-0.171*** (0.043)	-0.006 (0.010)	0.079*** (0.014)
Mode: Online	0.081 (0.162)	0.180*** (0.043)	0.069*** (0.010)	-0.026 (0.014)
Wordsum Score	0.931*** (0.275)	0.401*** (0.079)	0.103*** (0.019)	0.017 (0.025)
Word Count (log)	-1.016 (0.802)	0.659** (0.220)	0.009 (0.053)	-0.089 (0.070)
Constant	-3.972*** (0.586)	0.017 (0.165)	0.228*** (0.039)	0.365*** (0.052)
Observations	3,486	3,039	3,037	3,038
R ²		0.065	0.147	0.043
Log Likelihood	-1,084.839			

Note:

*p<0.05; **p<0.01; ***p<0.001

References

- Denny, Matthew J., and Arthur Spirling. 2018. "Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It." *Political Analysis* forthcoming.
- Lee, Moontae, and David Mimno. 2014. "Low-dimensional embeddings for interpretable anchor-based topic inference." In *Proceedings of Empirical Methods in Natural Language Processing*. Citeseer.
- Manning, Christopher D., Prabhakar Raghavan, Hinrich Schütze et al. 2008. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- Roberts, Margaret E, Brandon M Stewart, and Dustin Tingley. 2014. "stm: R package for structural topic models." *Journal of Statistical Software* 1: 1–49.
- Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58 (4): 1064–1082.