Women Also Know Stuff

Challenging the Gender Gap in Political Sophistication

Patrick W. Kraft[*]

Supplementary Material

[*]Ramón y Cajal Fellow; Juan March Institute and Department of Social Sciences, University Carlos III of Madrid, Spain. Email: patrickwilli.kraft@uc3m.es; ORCID iD: 0000-0003-0123-221X.

A   Data Overview

I   Open-Ended Items

2018 Cooperative Election Study (CES):   The main analysis focuses on a national stratified survey of 1,000 respondents as part of the 2018 CES. Among other items, the study includes the following open-ended questions:

- On the issue of **gun legislation**, please outline the main arguments that come to mind *in favor and against* background checks for all gun sales, including at gun shows and over the Internet.

- On the issue of **abortion**, please outline the main arguments that come to mind *in favor and against* banning abortions after the 20th week of pregnancy.

- On the issue of **immigration**, please outline the main arguments that come to mind *in favor and against* providing a legal status for recipients of the Deferred Action for Childhood Arrivals (DACA) status.

- On the issue of **health cure**, please outline the main arguments that come to mind *in favor and against* repealing the Affordable Care Act (Obamacare).

- On the issue of **trade policies**, please outline the main arguments that come to mind *in favor and against* imposing tariffs on imported steel and aluminum from countries including Canada, Europe, and Mexico.

2020, 2016, & 2012 American National Election Study (ANES):   In addition, I apply the measurement approach using three waves of the American National Election Study (ANES), each consisting of a representative survey of about 5000 adults in the months before the US Presidential election in each year. Here, discursive sophistication is evaluated using a set of 8 open-ended questions in which respondents were asked to list anything in particular that they like/dislike about the Democratic/Republican party as well as anything that might make them vote/not vote for either of the Presidential candidates.

2015 YouGov Survey   In order to replicate and extend the main analysis, I rely on a separate nationally representative survey employing yet another alternative set of open-ended responses. The data was collected by YouGov in December 2015 and contains responses of 1000 U.S. residents.[1]   As part of this study, respondents were asked to describe their attitudes towards two prominent political issues that were discussed frequently in the media. First, they were asked in a closed format whether they favor or oppose stricter gun laws. Subsequently, they were asked to respond to the following two questions:

- Still thinking about the question you just answered, what thoughts came to mind while you were answering that question? Please try to list everything that came to mind.

---

[1]See Clifford and Jerit (2018) for details on the study.

- Thinking about the mass shootings that have occurred in the U.S. in the last few years, what factors do you think are responsible for the shootings?

Second, the respondents reported on their attitudes towards the Affordable Care Act in a closed format and were then asked to elaborate in their own words by answering the following questions:

- Still thinking about the question you just answered, what thoughts came to mind while you were answering that question? Please try to list everything that came to mind.

- For decades, experts have observed that the United States spends far more per person on health care than any other country. However, the U.S. falls behind on most measures of health care outcomes, such as life expectancy. What factors do you think are responsible for the state of our health care system?

Swiss Referendum Survey Lastly, I examine survey data on Swiss citizens justifying their vote choices on multiple referenda used in a recent analysis by Colombo (2018). The author compiled a data set of cross-sectional surveys administered in Switzerland after national popular votes on multiple policy propositions. The original surveys were conducted as representative samples after each of thirty-four national policy votes that were held between 2008 and 2012 resulting in a total of about 27,000 observations. Respondents who participated in a given referendum (ca. 22,000 in total) were asked to describe the main reason as well as additional justifications for their decision in two separate items.

II   Conventional Knowledge Items

- *2018 CES*: Additive index containing 5 items (gun legislation, trade policy, DACA, health care, abortion).

- *2020 ANES*: Additive index containing 4 items (length of Senate term, federal government spending, majority in House, majority in Senate).

- *2016 ANES*: Additive index containing 4 items (length of Senate term, federal government spending, majority in House, majority in Senate).

- *2012 ANES*: Additive index containing 5 items (number of Presidential terms, size of budget deficit, length of Senate term, meaning of Medicare, federal government spending).

- *2015 YouGov*: Additive index containing 8 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).

III Ethical Standards

The human subjects research in this article was reviewed and approved as exempt by the University of Wisconsin-Milwaukee (UWM IRB# 20.044). The remaining research relies on secondary analyses of publicly available survey data such as the American National Election Studies (2012, 2016, 2020), the Cooperative Election Study (Schaffner, Ansolabehere, and Luks, 2019), as well as Clifford and Jerit (2018) and Colombo (2018).

B Detailed Information on Open-Ended Responses and Discursive Sophistication Components

I Distribution of Word Counts & Proportion of Non-Response
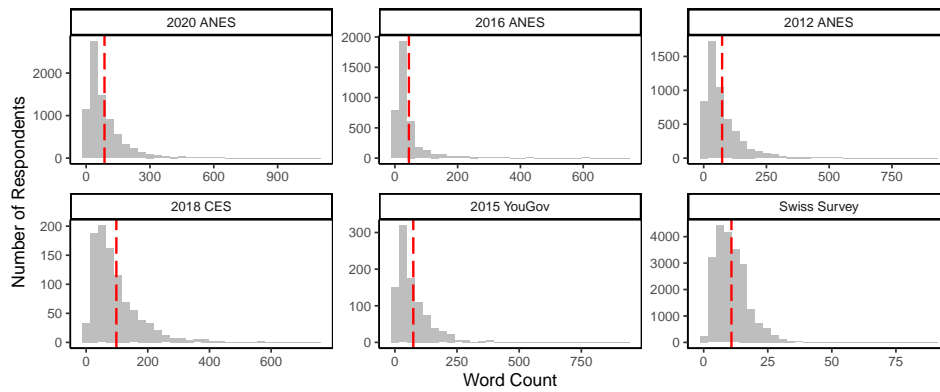
Figure B.1: Total word count across all open-ended responses for each survey participant. The dashed red lines indicate the average response lengths in each survey.
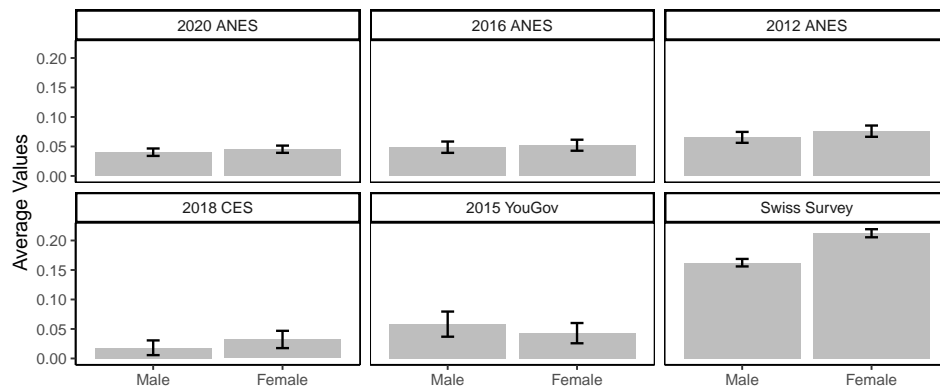
Figure B.2: Proportion of non-response comparing male and female survey participants (including 95% confidence intervals). Gender differences are only significant ($p < .05$) for the Swiss survey. Note, however, that respondents in the Swiss survey were only asked open-ended questions if they voted in the respective referendum.

II Preprocessing and Topic Model Specification

I rely on the structural topic model framework to extract and differentiate considerations mentioned by respondents. I follow the guidelines in Roberts et al. (2014) to preprocess our open-ended responses (lowercasing and stemming as well as removing stopwords, punctuation, numbers, and infrequent terms)[2] and used age, gender, education, party identification, as well as an interaction between education and party identification as covariates for topic prevalence. With the exception of gender, this variable selection is equivalent to the procedure described in Roberts et al. (2014) for open-ended survey responses such as those included in the ANES. I set the number of topics to 25 in order to focus on the differentiation of broader considerations, but I replicate equivalent results with larger numbers of topics below. Figure B.3 displays the resulting topic proportions for all data sets included in the analyses along with the most frequent and exclusive (FREX) terms associated with each topic.
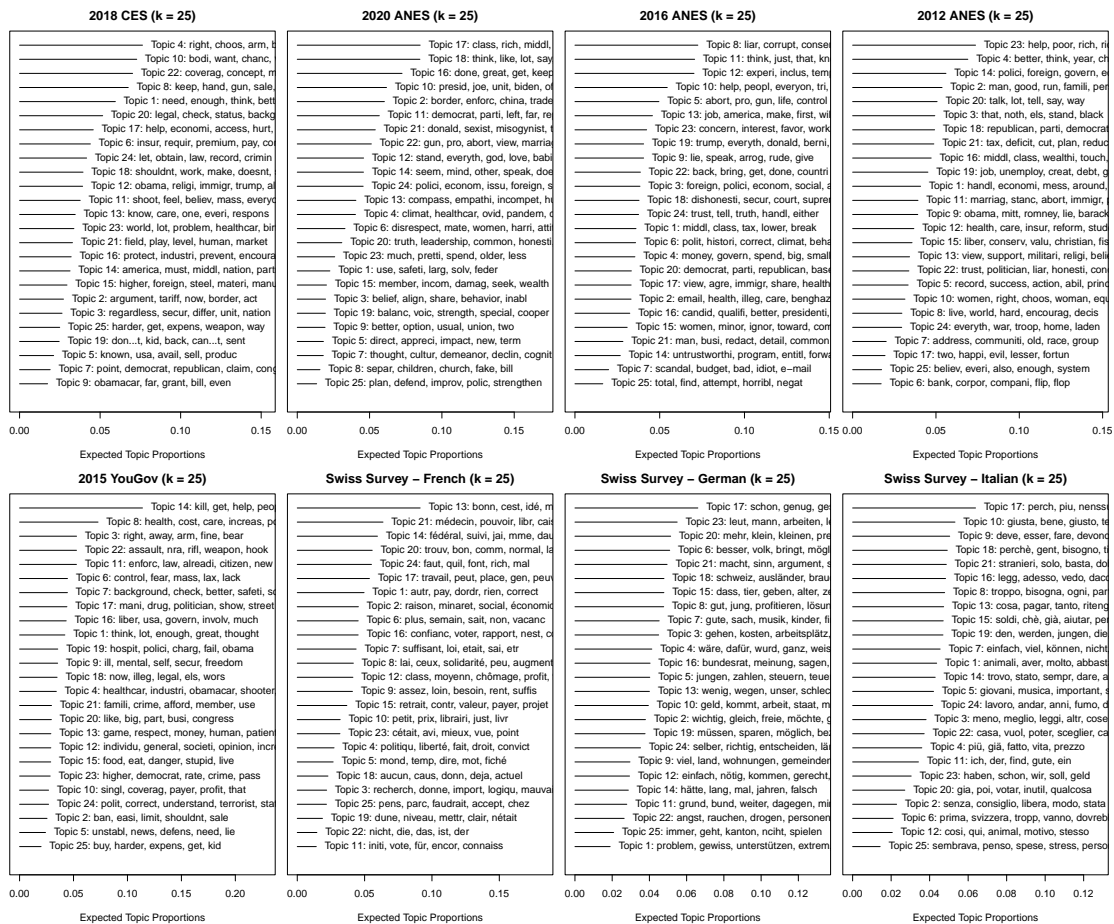
Figure B.3: Estimated topic proportions based on the structural topic model. See Appendix C.I for details on the model specification.

---

[2]Prior to applying these preprocessing steps, responses are cleaned by removing open-ended item non-response such as 'don't know' and correcting spelling errors using an implementation of the Aspell spell-checking algorithm (www.aspell.net).

III Discursive Sophistication Components
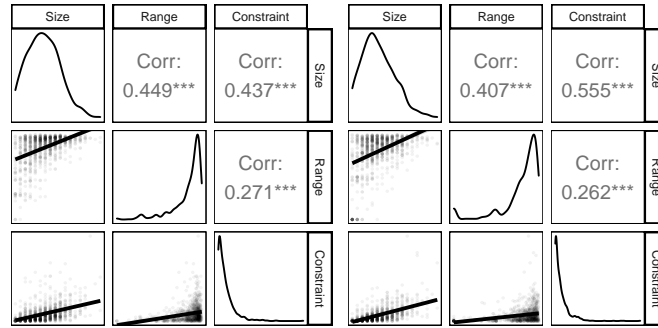
(a) 2018 CES   (b) 2015 YouGov
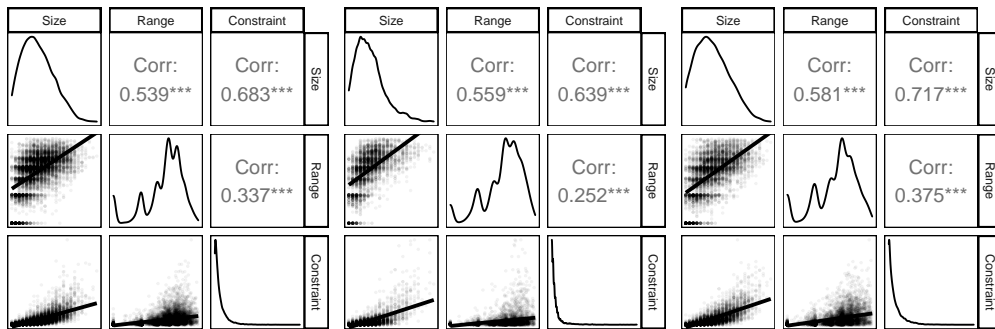
(c) 2020 ANES   (d) 2016 ANES   (e) 2012 ANES
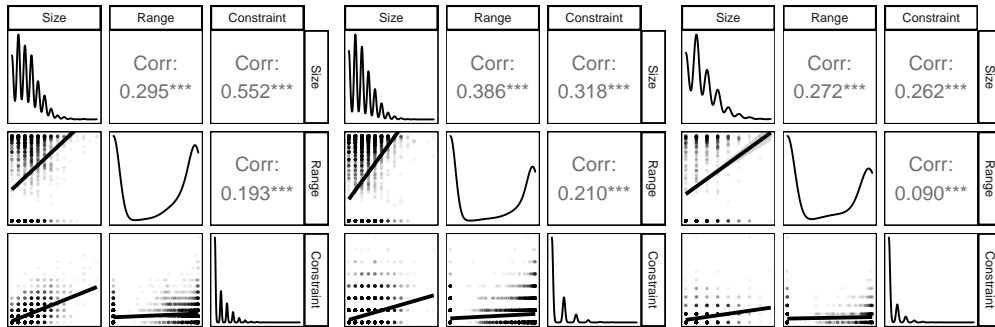
(f) Swiss Survey - French (g) Swiss Survey - German (h) Swiss Survey - Italian

Figure B.4: 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.

C Robustness Checks

I PreText Analysis

The first component of discursive sophistication (*size*) relies 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 preprocess the raw text and convert it into a document term matrix (DTM, see for example Manning et al., 2008). Common preprocessing 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 preprocessing choices. To address this issue, Denny and Spirling (2018) recommend that researchers compare DTMs under all possible preprocessing regimes. The authors propose *preText scores* as a measure to quantify the extent to which varying preprocessing regimes may yield unusual results compared to a baseline without any preprocessing. Following the procedure outlined in Denny and Spirling (2018), Figure C.1 displays the results of a linear model regressing preText scores resulting from all possible preprocessing regimes on each individual step for a random subset of 500 open-ended responses in each of the surveys included in the analyses. Significant coefficients indicate that the topic model results may be sensitive to the respective preprocessing step.
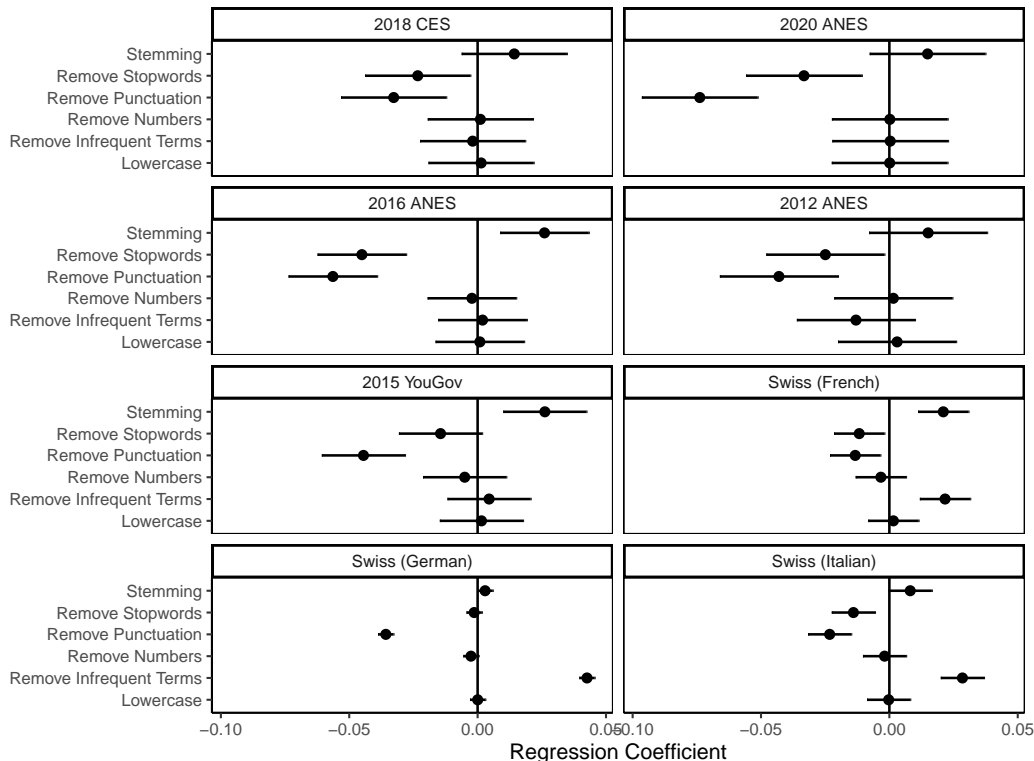
Figure C.1: PreText analysis of preprocessing decisions of open-ended responses across all datasets. Regression coefficients display the effects of each of the six preprocessing choices on the resulting preText score.

II   Discursive Sophistication for Varying Model Specifications

According to the analysis in Figure C.1, our results may be particularly sensitive to stemming and the removal of stopwords or punctuation. Denny and Spirling (2018), however, emphasize that the most important consideration in choosing preprocessing steps are theoretical. Given that the purpose of our topic model is to extract considerations related to political preferences, there are no theoretical reasons to incorporate punctuation since it does not contain any relevant content. It is less obvious from a theoretical perspective whether to use stemming or to remove stopwords from our open-ended responses, although it might be preferable in order to increase computational efficiency. Following Denny and Spirling (2018), I proceed by assessing to what extent discursive sophistication varies across the alternative preprocessing regimes identified as potentially influential. In addition, I consider another crucial modeling choice when working with topic models: determining the total number of topics k to be estimated.

Figure C.2 examines whether the proposed measure of discursive sophistication is sensitive to changing the number of topics k, stemming, and the removal of stopwords. The y-axis depicts the preferred preprocessing regime including all steps discussed above while the x-axis plots discursive sophistication resulting from alternative specifications. The panels on the left compare the preferred specification to discursive sophistication based on a larger number of topics ($k = 35$). The center panels does not use stemming as part of the preprocessing. The panels on the right do not remove stopwords prior to estimating the topic model. Across all data sets, discursive sophistication scores are highly correlated and therefore insensitive to preprocessing choices. Thus, the substantive results discussed in the main text are robust for alternative preprocessing regimes or varying numbers of topics.
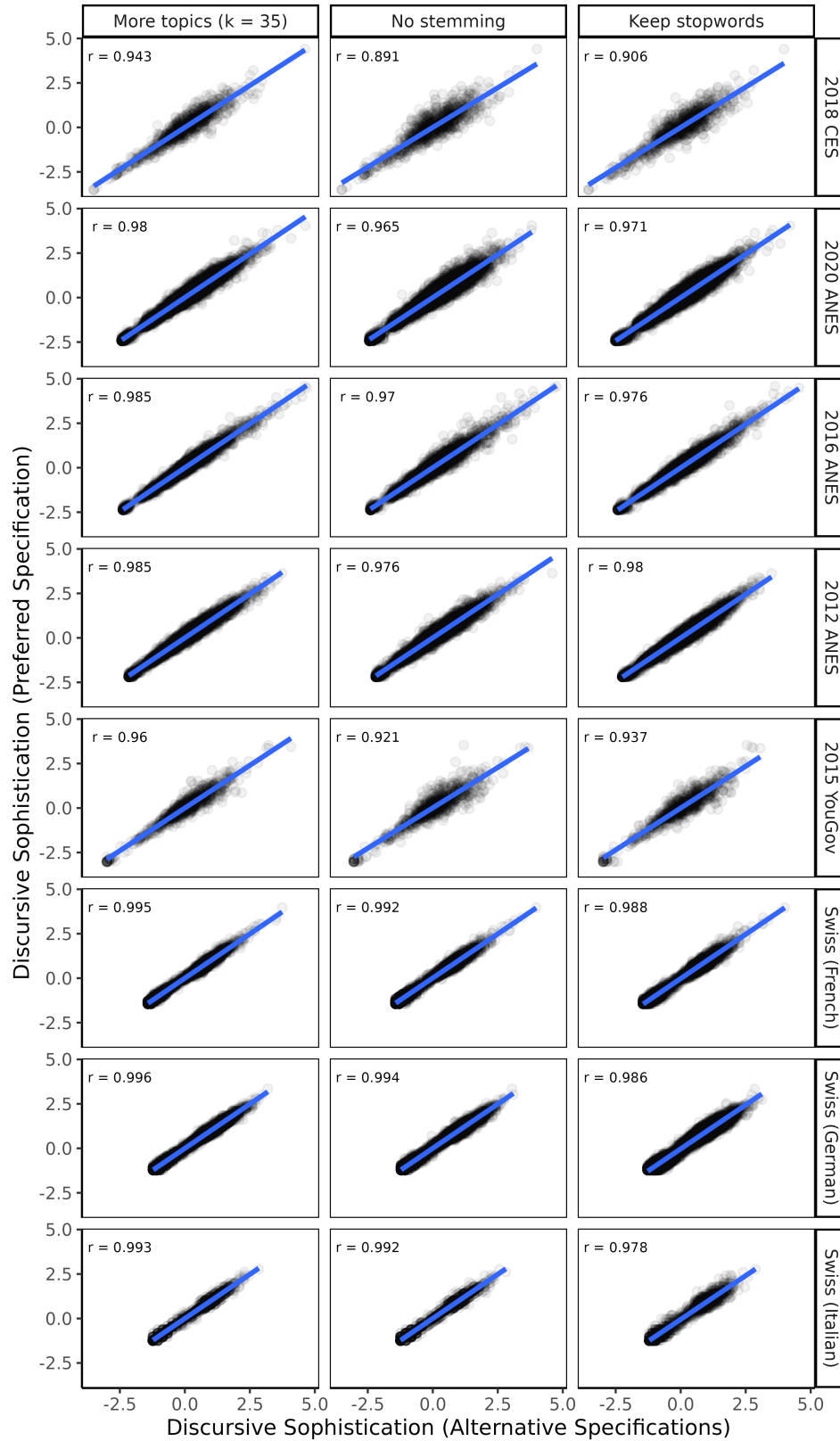
Figure C.2: Robustness of discursive sophistication measure for different preprocessing choices and topic model specifications.

III Controlling for Personality and Verbal Skills
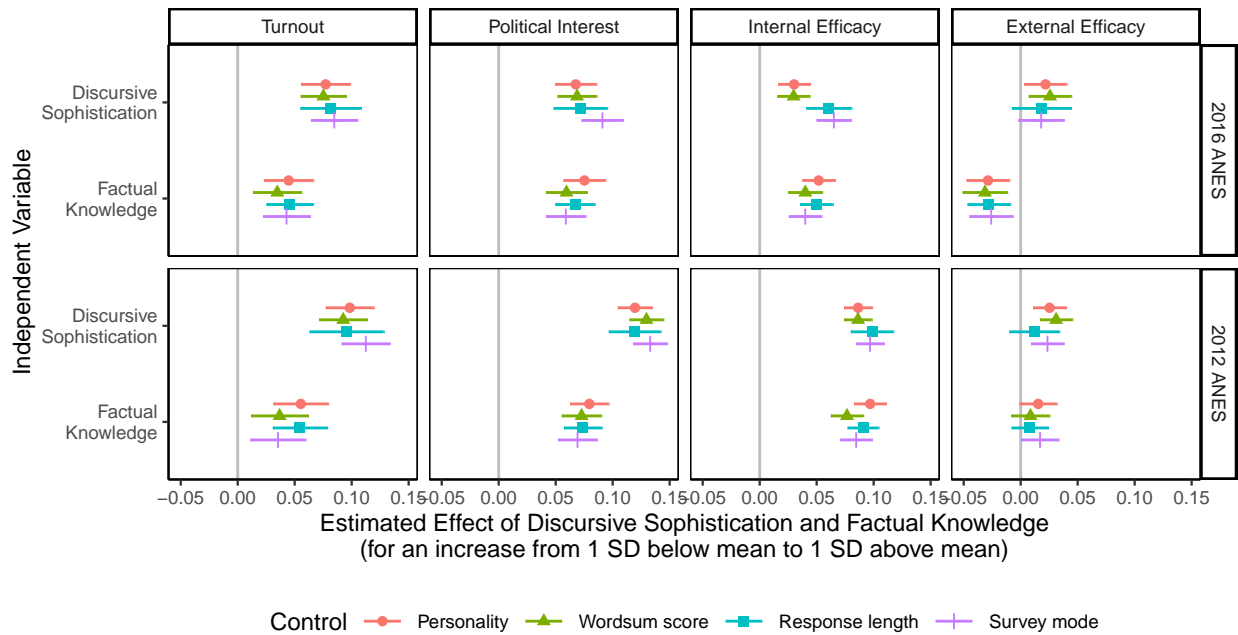
Figure C.3: Effects of sophistication on internal efficacy, external efficacy, non-conventional participation, and turnout in the 2012 and 2016 ANES. For each dependent variable, the figure displays the average marginal effects (AME) for each sophistication measure (including 95% confidence intervals). Model estimates are based on logistic regression (turnout) or OLS (political interest, internal efficacy, external efficacy). Compared to the specification used in Figure 2 in the main text (controlling for gender, education, income, age, race, and church attendance), the models displayed here include additional controls for personality (extraversion, openness to experience, being reserved), survey mode (online vs. face-to-face), verbal skills (Wordsum score), and overall verbosity (response length).

Description of additional control variables:

- *Personality*: Components of the Ten Item Personality Inventory (TIPI) measuring the "Big Five" personality traits measuring extraversion, openness to experience, being reserved.

- *Survey Mode*: Dichotomous indicator for face-to-face vs. online samples of the ANES surveys.

- *Wordsum vocabulary scores*: Modified version of the GSS wordsum vocabulary test consisting of 10 terms.

- *Response length*: Total number of words in the collection of open-ended responses by each individual.

9

IV   Discursive Sophistication and Uncertainty in Ideological Placements
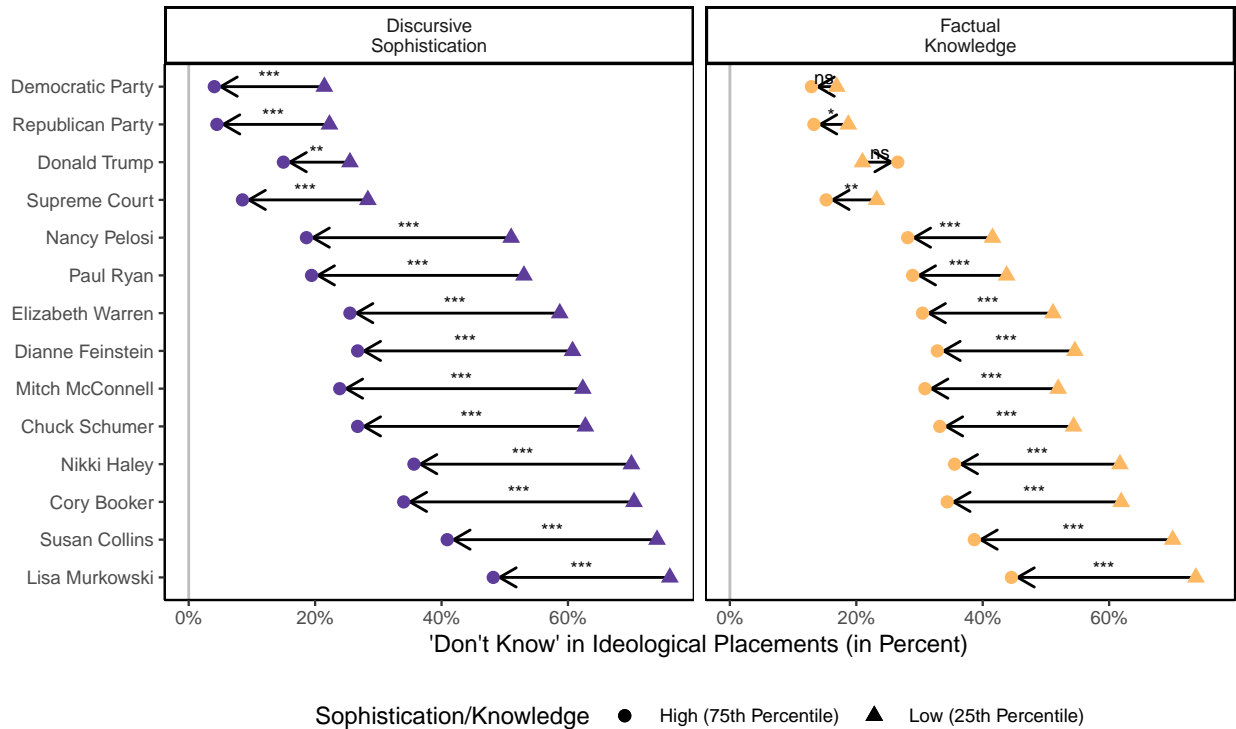
Figure C.4: Effects of sophistication on the probability of respondents to answer 'Don't Know' when asked to place individual politicians, parties, and institutions on the ideological spectrum from liberal to conservative in the 2018 CES. The figure shows the expected change comparing low (25th percentile) and high (75th percentile) levels of discursive sophistication / factual knowledge. Differences are statistically significant at: *p<0.05; **p<0.01; ***p<0.001.
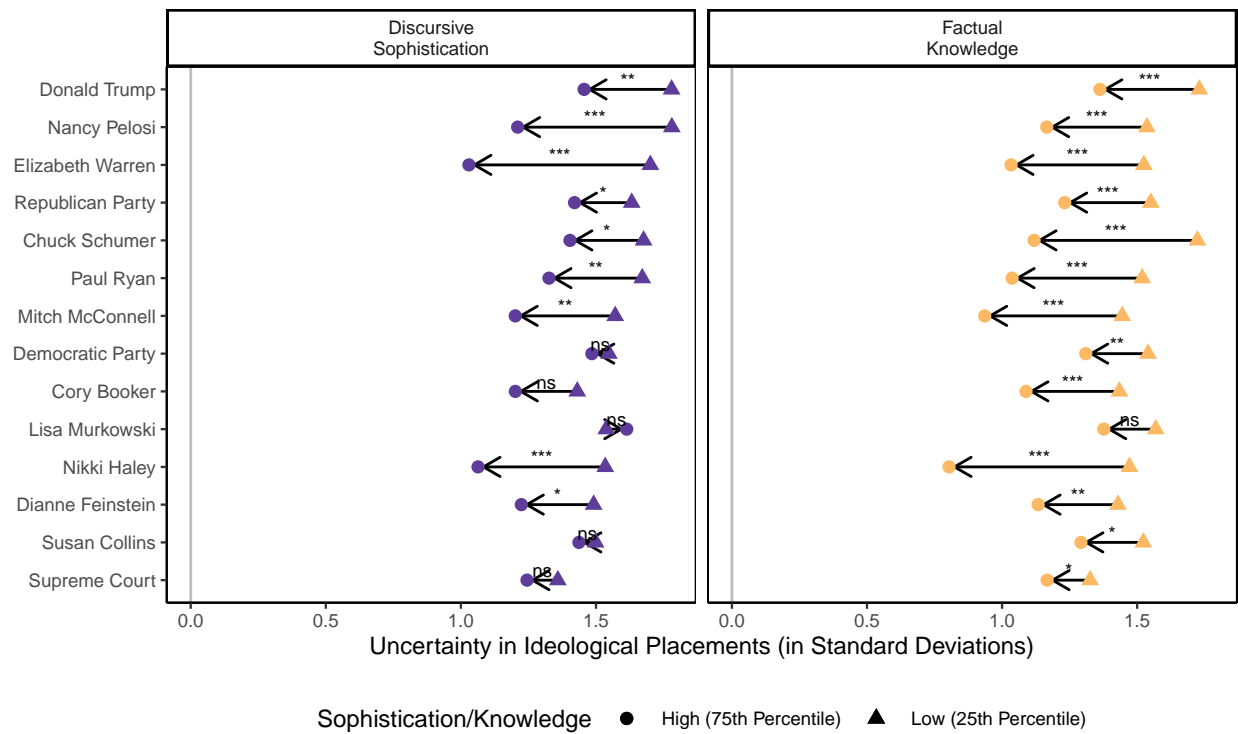
Figure C.5: Effects of sophistication on the uncertainty around ideological placements of politicians, parties, and institutions in the 2018 CES (measured in standard deviations). The figure shows the expected change comparing low (25th percentile) and high (75th percentile) levels of discursive sophistication / factual knowledge. Differences are statistically significant at: *p<0.05; **p<0.01; ***p<0.001.

V Discursive Sophistication and Ideological Proximity Voting
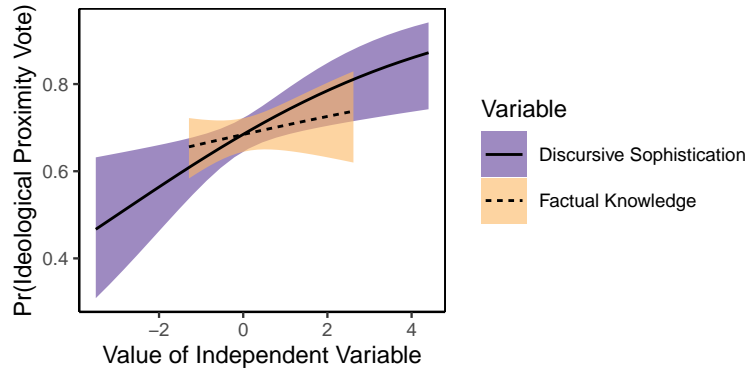
Figure C.6: Expected probability to vote for the senatorial candidate based on ideological proximity as a function of political sophistication (including 95% confidence intervals). Estimates are based on a logistic regression including controls for sociodemographic variables. The predictions are made by setting covariates equal to their mean (continuous) or median (categorical) value. Full regression results are displayed in Table C.1.

Table C.1: Logistic regression predicting ideological proximity-based voting for US Senators in the 2018 CES. Standard errors in parentheses. Estimates are used for Figure C.6.

	Dependent variable:
	Ideological Proximity Vote
Discursive Soph.	0.259**
	(0.096)
Factual Knowledge	0.100
	(0.100)
Female	−0.228
	(0.188)
Age	0.017**
	(0.005)
Black	−0.187
	(0.327)
College Degree	0.423*
	(0.199)
Household Income	0.318
	(0.441)
Church Attendance	0.107
	(0.273)
Constant	−0.227
	(0.343)
Observations	611
Akaike Inf. Crit.	748.749

Note: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

VI   Discursive Sophistication and Trust in Different News Sources
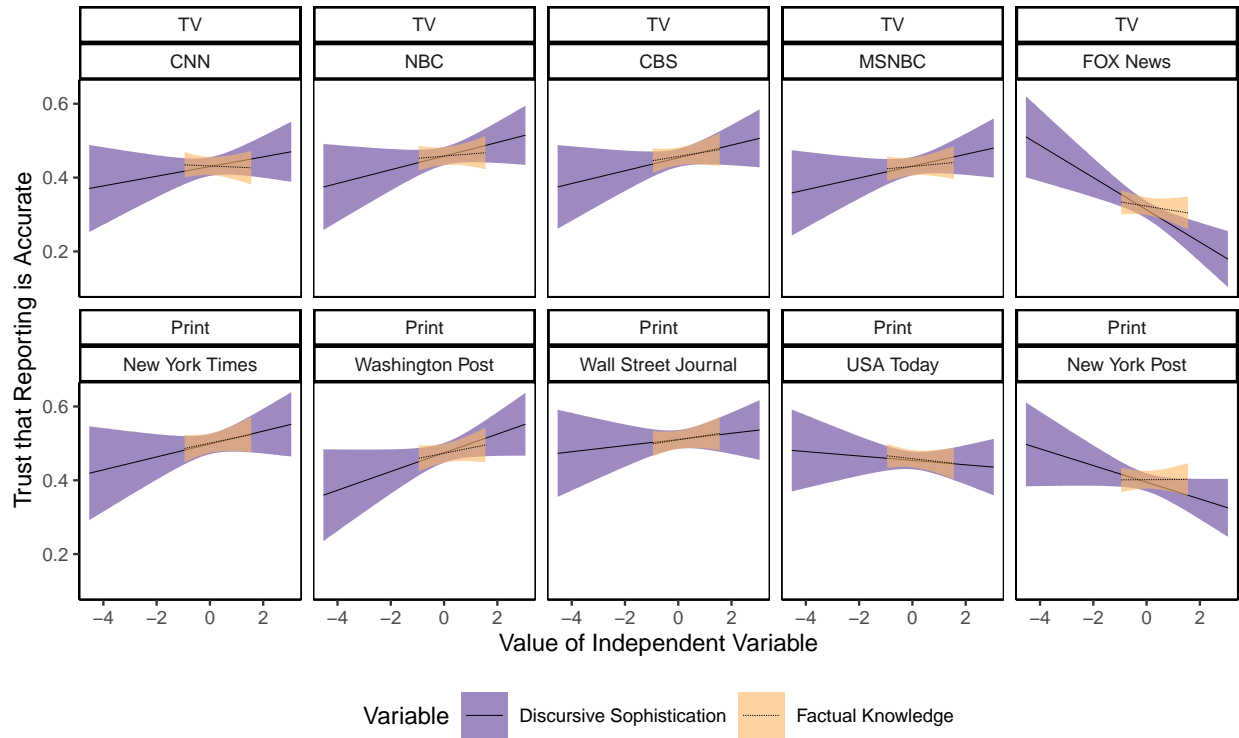
Figure C.7: Average trust in different news outlets as a function of political sophistication (including 95% confidence intervals). Estimates are based on bivariate linear regressions.

VII Personality and Verbal Skills as Predictors of Discursive Sophistication

Table C.2: Personality, verbal skills, and survey mode as predictors of discursive sophistication in the 2016 and 2012 ANES.

	Dependent variable:			
	Discursive Sophistication			
	2016 ANES		2012 ANES	
	(1)	(2)	(3)	(4)
Personality: Extraversion	0.060	0.062	−0.075	−0.069
	(0.069)	(0.069)	(0.061)	(0.060)
Personality: Openness to Experience	0.193*	0.190*	0.271***	0.288***
	(0.079)	(0.079)	(0.064)	(0.064)
Personality: Reserved	−0.101	−0.097	−0.126*	−0.129**
	(0.058)	(0.058)	(0.050)	(0.050)
Verbal Skills (Wordsum score)	0.953***	0.884***	1.110***	0.910***
	(0.082)	(0.084)	(0.070)	(0.072)
Survey Mode (Online)	−0.901***	−0.916***	−0.428***	−0.462***
	(0.035)	(0.035)	(0.031)	(0.031)
Factual Knowledge		0.064***		0.171***
		(0.017)		(0.017)
Female	−0.065*	−0.056	−0.033	0.008
	(0.032)	(0.032)	(0.027)	(0.027)
Age	0.001	0.0002	0.005***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Black	−0.061	−0.067	0.088*	0.116**
	(0.059)	(0.059)	(0.039)	(0.039)
PID: Democrat	0.036	0.029	0.046	0.052
	(0.039)	(0.038)	(0.032)	(0.032)
PID: Republican	0.053	0.052	0.032	0.022
	(0.041)	(0.041)	(0.036)	(0.036)
Education: High School	0.139	0.134	0.105	0.068
	(0.079)	(0.079)	(0.054)	(0.054)
Education: Some College	0.348***	0.341***	0.315***	0.259***
	(0.076)	(0.075)	(0.054)	(0.053)
Education: Bachelor's Degree	0.545***	0.535***	0.509***	0.424***
	(0.080)	(0.080)	(0.060)	(0.060)
Education: Graduate Degree	0.586***	0.573***	0.587***	0.487***
	(0.085)	(0.085)	(0.066)	(0.066)
Household Income	0.150**	0.139**	0.216***	0.156**
	(0.047)	(0.047)	(0.050)	(0.050)
Church Attendance	0.018	0.016	0.086*	0.087*
	(0.048)	(0.048)	(0.039)	(0.039)
Constant	−0.647***	−0.559***	−1.243***	−0.978***
	(0.120)	(0.122)	(0.095)	(0.097)
Observations	3,022	3,022	4,695	4,695
R^2	0.267	0.271	0.180	0.198

Note: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Table C.3: Personality, verbal skills, and survey mode as predictors of factual knowledge in the 2016 and 2012 ANES.

	Dependent variable:			
	Factual Knowledge			
	2016 ANES		2012 ANES	
	(1)	(2)	(3)	(4)
Personality: Extraversion	−0.040	−0.045	−0.040	−0.031
	(0.073)	(0.073)	(0.053)	(0.052)
Personality: Openness to Experience	0.054	0.040	−0.099	−0.134*
	(0.083)	(0.083)	(0.056)	(0.055)
Personality: Reserved	−0.050	−0.043	0.018	0.034
	(0.061)	(0.061)	(0.044)	(0.043)
Verbal Skills (Wordsum score)	1.083***	1.015***	1.166***	1.024***
	(0.086)	(0.088)	(0.061)	(0.062)
Survey Mode (Online)	0.229***	0.293***	0.197***	0.252***
	(0.037)	(0.041)	(0.027)	(0.027)
Discursive Soph.		0.071***		0.128***
		(0.019)		(0.013)
Female	−0.144***	−0.139***	−0.236***	−0.232***
	(0.034)	(0.034)	(0.023)	(0.023)
Age	0.006***	0.006***	0.008***	0.008***
	(0.001)	(0.001)	(0.001)	(0.001)
Black	0.088	0.093	−0.161***	−0.173***
	(0.062)	(0.062)	(0.034)	(0.034)
PID: Democrat	0.111**	0.109**	−0.033	−0.039
	(0.041)	(0.041)	(0.028)	(0.027)
PID: Republican	0.016	0.012	0.057	0.053
	(0.043)	(0.043)	(0.031)	(0.031)
Education: High School	0.088	0.078	0.211***	0.198***
	(0.083)	(0.083)	(0.047)	(0.046)
Education: Some College	0.103	0.078	0.331***	0.291***
	(0.080)	(0.080)	(0.046)	(0.046)
Education: Bachelor's Degree	0.153	0.114	0.498***	0.433***
	(0.085)	(0.085)	(0.052)	(0.052)
Education: Graduate Degree	0.197*	0.155	0.583***	0.508***
	(0.090)	(0.090)	(0.057)	(0.057)
Household Income	0.166***	0.155**	0.355***	0.328***
	(0.049)	(0.049)	(0.043)	(0.043)
Church Attendance	0.034	0.032	−0.003	−0.014
	(0.051)	(0.051)	(0.034)	(0.034)
Constant	−1.375***	−1.329***	−1.551***	−1.392***
	(0.127)	(0.127)	(0.082)	(0.082)
Observations	3,022	3,022	4,695	4,695
R^2	0.144	0.147	0.344	0.359

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

D Tables of Model Estimates

I Effects of sophistication on turnout, political interest, internal efficacy, and external efficacy (estimates used for Figure 2 in the main text)

Table D.1: Effects of sophistication on turnout and political interest in the 2018 CES. Standard errors in parentheses. Estimates of model (2) and (5) are used for Figure 2 in the main text.

	Dependent variable:					
	Turnout			Political Interest		
	(1)	(2)	(3)	(4)	(5)	(6)
Discursive Soph.	0.390***	0.253**	0.231*	0.047***	0.030***	0.030***
	(0.081)	(0.091)	(0.093)	(0.009)	(0.009)	(0.009)
Factual Knowledge	0.441***	0.292**	0.294**	0.083***	0.067***	0.068***
	(0.090)	(0.103)	(0.103)	(0.009)	(0.009)	(0.009)
Disc. X Factual			−0.107			−0.007
			(0.093)			(0.008)
Female		−0.516**	−0.520**		−0.081***	−0.080***
		(0.189)	(0.190)		(0.017)	(0.017)
Age		0.044***	0.044***		0.004***	0.004***
		(0.006)	(0.006)		(0.0005)	(0.0005)
Black		−0.592*	−0.601*		−0.041	−0.041
		(0.266)	(0.267)		(0.028)	(0.028)
College Degree		0.651**	0.671**		0.033	0.034
		(0.206)	(0.207)		(0.018)	(0.018)
Household Income		1.086*	1.074*		0.159***	0.157***
		(0.461)	(0.461)		(0.041)	(0.042)
Church Attendance		0.663*	0.655*		0.060*	0.060*
		(0.281)	(0.282)		(0.025)	(0.025)
Constant	1.204***	−1.205***	−1.177***	0.623***	0.397***	0.400***
	(0.082)	(0.339)	(0.340)	(0.008)	(0.032)	(0.032)
Observations	943	849	849	941	848	848
R^2				0.149	0.276	0.277
Akaike Inf. Crit.	994.156	795.110	795.804			

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

Table D.2: Effects of sophistication on turnout, political interest, internal efficacy, and external efficacy in the 2018 CES. Standard errors in parentheses. Estimates of model (2) and (5) are used for Figure 2 in the main text.

	Dependent variable:					
	Internal Efficacy			External Efficacy		
	(1)	(2)	(3)	(4)	(5)	(6)
Discursive Soph.	0.047***	0.036***	0.036***	−0.004	−0.011	−0.010
	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)
Factual Knowledge	0.073***	0.062***	0.063***	0.018*	0.019	0.017
	(0.008)	(0.008)	(0.008)	(0.009)	(0.010)	(0.010)
Disc. X Factual			−0.003			0.011
			(0.007)			(0.009)
Female		−0.050**	−0.050**		0.020	0.020
		(0.016)	(0.016)		(0.018)	(0.018)
Age		0.002***	0.002***		−0.0004	−0.0004
		(0.0004)	(0.0004)		(0.001)	(0.001)
Black		0.030	0.030		−0.026	−0.026
		(0.026)	(0.026)		(0.030)	(0.030)
College Degree		0.040*	0.041*		0.037	0.036
		(0.017)	(0.017)		(0.020)	(0.020)
Household Income		0.115**	0.114**		0.027	0.029
		(0.038)	(0.038)		(0.044)	(0.044)
Church Attendance		0.007	0.007		0.086**	0.086**
		(0.023)	(0.023)		(0.026)	(0.026)
Constant	0.607***	0.455***	0.456***	0.342***	0.299***	0.294***
	(0.007)	(0.029)	(0.029)	(0.008)	(0.034)	(0.034)
Observations	941	847	847	941	848	848
R^2	0.159	0.222	0.222	0.004	0.025	0.027

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

Table D.3: Effects of sophistication on turnout and political interest in the 2020 ANES. Standard errors in parentheses. Estimates of model (2) and (5) are used for Figure 2 in the main text.

	Dependent variable:					
	Turnout			Political Interest		
	(1)	(2)	(3)	(4)	(5)	(6)
Discursive Soph.	0.519***	0.379***	0.375***	0.054***	0.052***	0.052***
	(0.039)	(0.042)	(0.045)	(0.003)	(0.003)	(0.003)
Factual Knowledge	0.400***	0.274***	0.271***	0.059***	0.041***	0.041***
	(0.041)	(0.047)	(0.049)	(0.003)	(0.003)	(0.003)
Disc. X Factual			−0.010			−0.003
			(0.044)			(0.003)
Female		0.445***	0.444***		−0.042***	−0.042***
		(0.085)	(0.085)		(0.006)	(0.006)
Age		0.030***	0.030***		0.003***	0.003***
		(0.003)	(0.003)		(0.0002)	(0.0002)
Black		0.247	0.247		0.021*	0.021*
		(0.137)	(0.137)		(0.010)	(0.010)
College Degree		0.704***	0.706***		0.014*	0.015*
		(0.100)	(0.100)		(0.006)	(0.006)
Household Income		1.298***	1.299***		0.016	0.016
		(0.142)	(0.142)		(0.009)	(0.009)
Church Attendance		0.565***	0.565***		−0.012	−0.012
		(0.143)	(0.143)		(0.009)	(0.009)
Constant	2.162***	−0.634***	−0.635***	0.680***	0.509***	0.510***
	(0.043)	(0.154)	(0.154)	(0.003)	(0.011)	(0.011)
Observations	6,695	6,371	6,371	7,324	6,965	6,965
R^2				0.127	0.186	0.186
Akaike Inf. Crit.	4,554.946	3,974.159	3,976.112			

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

18

Table D.4: Effects of sophistication on internal and external efficacy in the 2020 ANES. Standard errors in parentheses. Estimates of model (2) and (5) are used for Figure 2 in the main text.

| | *Dependent variable:* | | | | | |
| | Turnout | | | Political Interest | | |
	(1)	(2)	(3)	(4)	(5)	(6)
Discursive Soph.	0.041***	0.037***	0.037***	0.016***	0.006	0.006
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Factual Knowledge	0.053***	0.042***	0.042***	0.024***	0.020***	0.019***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Disc. X Factual			0.007**			0.010**
			(0.002)			(0.003)
Female		−0.034***	−0.034***		0.028***	0.029***
		(0.005)	(0.005)		(0.007)	(0.007)
Age		0.001***	0.001***		−0.00003	−0.00002
		(0.0001)	(0.0001)		(0.0002)	(0.0002)
Black		0.054***	0.053***		0.029*	0.029*
		(0.009)	(0.009)		(0.012)	(0.012)
College Degree		0.033***	0.032***		0.066***	0.066***
		(0.005)	(0.005)		(0.007)	(0.007)
Household Income		0.039***	0.039***		0.056***	0.056***
		(0.008)	(0.008)		(0.011)	(0.011)
Church Attendance		−0.028***	−0.028***		0.014	0.015
		(0.007)	(0.007)		(0.010)	(0.010)
Constant	0.590***	0.529***	0.526***	0.355***	0.273***	0.269***
	(0.002)	(0.010)	(0.010)	(0.003)	(0.013)	(0.013)
Observations	6,636	6,319	6,319	6,633	6,316	6,316
R^2	0.136	0.161	0.162	0.015	0.042	0.044

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

Table D.5: Effects of sophistication on turnout and political interest in the 2016 ANES. Standard errors in parentheses. Estimates of model (2) and (5) are used for Figure 2 in the main text.

	Dependent variable:					
	Turnout			Political Interest		
	(1)	(2)	(3)	(4)	(5)	(6)
Discursive Soph.	0.542***	0.441***	0.440***	0.048***	0.037***	0.037***
	(0.055)	(0.058)	(0.060)	(0.004)	(0.004)	(0.004)
Factual Knowledge	0.333***	0.250***	0.250***	0.049***	0.033***	0.034***
	(0.053)	(0.057)	(0.061)	(0.004)	(0.004)	(0.004)
Disc. X Factual			−0.001			0.003
			(0.056)			(0.004)
Female		0.165	0.165		−0.065***	−0.065***
		(0.112)	(0.112)		(0.008)	(0.008)
Age		0.022***	0.022***		0.004***	0.004***
		(0.003)	(0.003)		(0.0002)	(0.0002)
Black		1.028***	1.028***		0.022	0.022
		(0.233)	(0.233)		(0.015)	(0.015)
College Degree		0.678***	0.678***		0.062***	0.062***
		(0.137)	(0.137)		(0.009)	(0.009)
Household Income		0.532***	0.532***		0.027*	0.027*
		(0.157)	(0.157)		(0.012)	(0.012)
Church Attendance		1.030***	1.030***		0.006	0.006
		(0.186)	(0.186)		(0.012)	(0.012)
Constant	2.183***	0.080	0.080	0.638***	0.440***	0.440***
	(0.058)	(0.201)	(0.201)	(0.004)	(0.017)	(0.017)
Observations	3,685	3,575	3,575	3,708	3,595	3,595
R^2				0.072	0.156	0.156
Akaike Inf. Crit.	2,479.933	2,269.075	2,271.075			

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

Table D.6: Effects of sophistication on internal and external efficacy in the 2016 ANES. Standard errors in parentheses. Estimates of model (2) and (5) are used for Figure 2 in the main text.

| | Dependent variable: | | | | | |
| | Internal Efficacy | | | External Efficacy | | |
	(1)	(2)	(3)	(4)	(5)	(6)
Discursive Soph.	0.029***	0.018***	0.019***	0.022***	0.012*	0.013**
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
Factual Knowledge	0.037***	0.026***	0.026***	−0.006	−0.014**	−0.014**
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
Disc. X Factual			0.001			0.006
			(0.004)			(0.005)
Female		−0.055***	−0.054***		−0.002	−0.001
		(0.007)	(0.007)		(0.009)	(0.009)
Age		0.001**	0.001**		−0.00000	−0.00001
		(0.0002)	(0.0002)		(0.0003)	(0.0003)
Black		0.042***	0.042***		−0.016	−0.017
		(0.012)	(0.012)		(0.016)	(0.016)
College Degree		0.071***	0.071***		0.059***	0.058***
		(0.008)	(0.008)		(0.010)	(0.010)
Household Income		0.033**	0.033**		0.046***	0.046***
		(0.010)	(0.010)		(0.013)	(0.013)
Church Attendance		−0.012	−0.012		0.081***	0.081***
		(0.010)	(0.010)		(0.013)	(0.013)
Constant	0.541***	0.484***	0.484***	0.397***	0.316***	0.314***
	(0.004)	(0.014)	(0.014)	(0.005)	(0.018)	(0.018)
Observations	3,201	3,114	3,114	3,203	3,116	3,116
R^2	0.057	0.114	0.114	0.008	0.040	0.041

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

Table D.7: Effects of sophistication on turnout and political interest in the 2012 ANES. Standard errors in parentheses. Estimates of model (2) and (5) are used for Figure 2 in the main text.

	Dependent variable:					
	Turnout			Political Interest		
	(1)	(2)	(3)	(4)	(5)	(6)
Discursive Soph.	0.492***	0.400***	0.409***	0.069***	0.065***	0.065***
	(0.041)	(0.043)	(0.044)	(0.004)	(0.004)	(0.004)
Factual Knowledge	0.378***	0.216***	0.232***	0.055***	0.037***	0.037***
	(0.041)	(0.048)	(0.050)	(0.004)	(0.004)	(0.004)
Disc. X Factual			0.049			0.009*
			(0.044)			(0.004)
Female		0.076	0.079		−0.064***	−0.064***
		(0.084)	(0.084)		(0.007)	(0.007)
Age		0.027***	0.027***		0.003***	0.003***
		(0.003)	(0.003)		(0.0002)	(0.0002)
Black		0.809***	0.807***		0.041***	0.041***
		(0.120)	(0.120)		(0.010)	(0.010)
College Degree		0.478***	0.468***		0.031***	0.029***
		(0.109)	(0.110)		(0.008)	(0.008)
Household Income		1.103***	1.106***		0.002	0.002
		(0.153)	(0.153)		(0.013)	(0.013)
Church Attendance		0.650***	0.652***		0.009	0.010
		(0.127)	(0.127)		(0.010)	(0.010)
Constant	1.660***	−0.604***	−0.613***	0.607***	0.477***	0.474***
	(0.042)	(0.148)	(0.148)	(0.004)	(0.014)	(0.014)
Observations	4,846	4,713	4,713	5,163	5,001	5,001
R^2				0.131	0.175	0.176
Akaike Inf. Crit.	4,193.841	3,799.204	3,799.996			

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

Table D.8: Effects of sophistication on internal and external efficacy in the 2012 ANES. Standard errors in parentheses. Estimates of model (2) and (5) are used for Figure 2 in the main text.

	Dependent variable:					
	Internal Efficacy			External Efficacy		
	(1)	(2)	(3)	(4)	(5)	(6)
Discursive Soph.	0.049***	0.046***	0.046***	0.021***	0.015***	0.015***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Factual Knowledge	0.052***	0.046***	0.046***	−0.002	0.004	0.004
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Disc. X Factual			0.006*			0.010**
			(0.003)			(0.004)
Female		−0.051***	−0.051***		0.017*	0.017*
		(0.006)	(0.006)		(0.007)	(0.007)
Age		0.00000	0.00001		−0.0003	−0.0003
		(0.0002)	(0.0002)		(0.0002)	(0.0002)
Black		0.054***	0.054***		0.084***	0.084***
		(0.008)	(0.008)		(0.009)	(0.009)
College Degree		0.031***	0.030***		0.035***	0.033***
		(0.007)	(0.007)		(0.008)	(0.008)
Household Income		0.022*	0.022*		0.012	0.012
		(0.011)	(0.011)		(0.013)	(0.013)
Church Attendance		0.005	0.005		0.049***	0.050***
		(0.008)	(0.008)		(0.010)	(0.010)
Constant	0.569***	0.564***	0.562***	0.391***	0.349***	0.345***
	(0.003)	(0.011)	(0.011)	(0.003)	(0.013)	(0.013)
Observations	5,153	4,993	4,993	5,139	4,982	4,982
R^2	0.129	0.155	0.156	0.007	0.037	0.039
Note:				*p<0.05; **p<0.01; ***p<0.001		

II Expected information retrieval in the 2015 YouGov Study as a function of political sophistication (estimates used for Figure 3 in the main text)

Table D.9: Linear regressions predicting information retrieval in the 2015 YouGov study. Standard errors in parentheses. Estimates of model (2) are used for Figure 3 in the main text.

	Dependent variable:		
	Information Retrieval		
	(1)	(2)	(3)
Discursive Soph.	0.362***	0.292***	0.285***
	(0.052)	(0.055)	(0.056)
Factual Knowledge	0.249***	0.315***	0.310***
	(0.053)	(0.065)	(0.065)
Disc. X Factual			−0.053
			(0.052)
Female		0.389***	0.387***
		(0.112)	(0.112)
Age		0.007*	0.006
		(0.003)	(0.003)
Black		−0.217	−0.206
		(0.191)	(0.191)
College Degree		0.158	0.156
		(0.126)	(0.126)
Household Income		−0.584*	−0.576*
		(0.270)	(0.270)
Church Attendance		−0.538***	−0.539***
		(0.150)	(0.149)
Constant	7.452***	7.277***	7.306***
	(0.049)	(0.193)	(0.195)
Observations	916	790	790
R^2	0.102	0.134	0.135
Note:	*p<0.05; **p<0.01; ***p<0.001		

III The gender gap in political sophistication after controlling for common determinants (estimates used for Figure 6 in the main text)

Table D.10: Linear regressions predicting discursive sophistication in the CES, ANES, and YouGov study. Estimates are used for Figure 6 in the main text.

	Dependent variable:				
	Discursive Sophistication				
	2018 CES	2020 ANES	2016 ANES	2012 ANES	2015 YouGov
	(1)	(2)	(3)	(4)	(5)
Female	−0.017	−0.058*	−0.034	−0.005	0.064
	(0.069)	(0.023)	(0.033)	(0.027)	(0.070)
Age	0.006**	0.004***	0.003***	0.005***	0.007***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Black	−0.342**	−0.280***	−0.130*	0.011	−0.547***
	(0.114)	(0.041)	(0.057)	(0.036)	(0.126)
College Degree	0.387***	0.444***	0.409***	0.455***	0.168*
	(0.073)	(0.024)	(0.035)	(0.031)	(0.083)
Household Income	0.121	0.443***	0.279***	0.408***	0.100
	(0.170)	(0.038)	(0.046)	(0.049)	(0.177)
Church Attendance	0.054	0.030	0.071	0.083*	−0.033
	(0.101)	(0.035)	(0.048)	(0.039)	(0.100)
Constant	−0.460***	−0.633***	−0.530***	−0.603***	−0.383**
	(0.129)	(0.045)	(0.063)	(0.051)	(0.127)
Observations	849	6,966	3,595	5,003	790
R^2	0.067	0.111	0.076	0.091	0.050

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

Table D.11: Linear regressions predicting discursive sophistication in the Swiss referendum study. Estimates are used for Figure 6 in the main text.

	Dependent variable:		
	Discursive Sophistication		
	French	German	Italian
	(1)	(2)	(3)
Female	0.019	−0.048**	0.034
	(0.029)	(0.018)	(0.048)
Age	0.002*	−0.001*	0.001
	(0.001)	(0.001)	(0.002)
College Degree	−0.096**	0.006	−0.013
	(0.029)	(0.018)	(0.048)
Constant	−0.065	0.093**	−0.086
	(0.058)	(0.035)	(0.097)
Observations	4,994	12,465	1,840
R^2	0.004	0.001	0.001

Note: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Table D.12: Linear regressions predicting factual knowledge in the CES, ANES, and YouGov study. Estimates are used for Figure 6 in the main text.

	Dependent variable:				
	Factual Knowledge				
	2018 CES	2020 ANES	2016 ANES	2012 ANES	2015 YouGov
	(1)	(2)	(3)	(4)	(5)
Female	−0.150*	−0.351***	−0.101**	−0.253***	−0.597***
	(0.067)	(0.022)	(0.032)	(0.024)	(0.060)
Age	0.003	0.012***	0.008***	0.012***	0.012***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Black	−0.284*	−0.173***	−0.036	−0.374***	−0.322**
	(0.111)	(0.039)	(0.056)	(0.032)	(0.108)
College Degree	0.508***	0.437***	0.213***	0.428***	0.413***
	(0.071)	(0.023)	(0.034)	(0.027)	(0.071)
Household Income	0.254	0.393***	0.344***	0.654***	0.852***
	(0.165)	(0.037)	(0.045)	(0.043)	(0.152)
Church Attendance	−0.320**	−0.019	−0.084	−0.044	−0.198*
	(0.098)	(0.034)	(0.047)	(0.034)	(0.085)
Constant	−0.193	−0.803***	−0.626***	−0.746***	−0.479***
	(0.125)	(0.043)	(0.062)	(0.045)	(0.109)
Observations	849	6,966	3,595	5,003	790
R^2	0.112	0.171	0.068	0.239	0.277

Note: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

References

American National Election Studies. 2012. "ANES 2012 Time Series Study Full Release [dataset and documentation].". https://www.electionstudies.org.

American National Election Studies. 2016. "ANES 2016 Time Series Study Full Release [dataset and documentation].". https://www.electionstudies.org.

American National Election Studies. 2020. "ANES 2020 Time Series Study Full Release [dataset and documentation].". July 19, 2021 version. https://www.electionstudies.org.

Clifford, Scott, and Jennifer Jerit. 2018. "Disgust, Anxiety, and Political Learning in the Face of Threat." *American Journal of Political Science* 62 (2): 266–279.

Colombo, Céline. 2018. "Justifications and Citizen Competence in Direct Democracy: A Multilevel Analysis." *British Journal of Political Science* 48 (3): 787–806.

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* 26 (2): 168–189.

Manning, Christopher D., Prabhakar Raghavan, Hinrich Schütze et al. 2008. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

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.

Schaffner, Brian, Stephen Ansolabehere, and Sam Luks. 2019. "CCES Common Content, 2018.". https://doi.org/10.7910/DVN/ZSBZ7K.