Looking for Answers
Measuring Political Sophistication in Open-Ended Responses[*]

Patrick W. Kraft[†]

July 19, 2017

– WORK IN PROGRESS –

PLEASE DO NOT CITE OR REDISTRIBUTE WITHOUT PERMISSION

Abstract

This paper proposes a general framework to assess political sophistication in verbatim responses to open-ended survey items. The measure aims to capture the elaborateness of individual attitude expressions by examining their relative length, topic diversity, and opinion diversity. I validate the approach by comparing the measure to conventional political knowledge metrics in two separate studies using different batteries of open-ended items. The paper proceeds to illustrate how text-based sophistication helps refine important previous insights from the literature, for example regarding the oft-cited gender gap in political knowledge.

Introduction

> *"Trump owes his victory to the uninformed. But it's not just Trump. Political scientists have been studying what voters know and how they think for well over 65 years. The results are frightening."*

Two days after the 2016 US Presidential election, Jason Brennan provided this— rather grim— view of the electorate in an article titled "Trump Won Because Voters Are Ignorant, Literally" published on *ForeignPolicy.com*. While the election of Donald J. Trump is certainly exceptional in various ways, the assessment of alarmingly high levels of voter ignorance can hardly be seen as a novel phenomenon. Quite contrary, it has been one of the major reoccurring themes in public opinion research. Not too long ago, for example, Bartels (2005) attributed public support for the Bush administration's 2001 and 2003 tax cuts to a substantial lack of political information among voters (but see Lupia et al., 2007; Bartels, 2007). Similarly, Delli Carpini and Keeter's (1996) seminal book on political knowledge warned that widespread ignorance might jeopardize equal representation of citizens. Early influential scholars such as Converse (1964) also emphasized that large parts of the public lack a sufficient understanding of abstract ideological concepts and do not hold stable issue positions. Indeed, the finding that citizens appear to know too little about politics seems to go as far back in history as the systematic study of political knowledge itself.

Yet, not everyone agrees with this pessimistic assessment. Instead, there has been a lively debate about how to accurately assess political knowledge in the first place (e.g. Mondak, 2000; Mondak and Davis, 2001; Sturgis, Allum, and Smith, 2008; DeBell, 2013; Pietryka and MacIntosh, 2013). Most analyses rely on standard item batteries that assess individuals' factual knowledge about political institutions and officeholders (e.g., Delli Carpini and Keeter, 1996). However, recent research points to important distinctions in types of political knowledge that have previously been disregarded (Barabas et al., 2014). Furthermore, scholars argue that factual knowledge does not necessarily capture how people structure their attitudes and beliefs (e.g. Luskin, 1987) and may not be theoretically relevant for political participation (Lupia, 2006).

The last point raises a fundamental question about the type of knowledge that is actually

necessary for outcomes we view as desirable (see also Lupia, 2015): Do citizens really need to know specific details about US political institutions in order to successfully engage in politics? For example in the context of elections, precise knowledge about the role of the Supreme Court in judicial review can hardly be viewed as necessary to cast an informed vote between two candidates. This is especially the case since voters can rely on heuristics (Lupia, 1994) or procedural knowledge (Prior and Lupia, 2008) when forming political preferences. Factual knowledge as measured in many public opinion surveys can therefore not be seen as an ideal indicator for citizens' competence in making informed vote choices.

The present paper proposes an alternative measure of political sophistication that aims to address this disconnect. Normative democratic theory suggests that voters should hold informed opinions about the candidate and issues before casting a vote. Rather than relying on factual knowledge that is potentially unrelated to the task at hand, I therefore examine how respondents discuss their preferences and beliefs about political candidates in their own words. For a given set of verbatim responses, the measure asses political sophistication based the relative response length, the diversity in topics raised by individuals, as well as their diversity in opinions. The goal is to assess whether political attitudes are expressed in a more elaborate manner — a question that is not directly discernible from factual knowledge items. The text-based measure is therefore conceptually closer to the degree of structure and constraint in political belief systems (see for example Tetlock, 1983; Luskin, 1987).

Overall, the measure has similar characteristics as conventional factual knowledge scores. Indeed, text-based sophistication is a stronger predictor of internal efficacy, political engagement, and turnout than most traditional measures. Contrary to previous research, however, there is no evidence for a gender gap in political sophistication. While women might score lower than men on factual knowledge about political institutions and elites, there are no differences in the complexity of expressed political attitudes. More generally, the results suggest that developing valid measures of political sophistication based on open-ended responses can provide new opportunities to examine political knowledge across time and contexts.

Political Knowledge and Sophistication

In his seminal study on political attitudes, Converse (1964) examined whether citizens hold constrained belief systems about politics. Belief systems are defined as "a configuration of ideas and attitudes in which the elements are bound together by some form of constraint or functional interdependence" (Converse, 1964, 207). Converse found that the majority of the electorate does not hold structured and constrained belief systems, understand abstract ideological concepts, or hold stable issue positions. This pessimistic view regarding the competence of the U.S. electorate has been supported in multiple subsequent analyses. Delli Carpini and Keeter (1996) show that large parts of the American electorate are not sufficiently informed about politics and that there are systematic differences in political attitudes and behavior between citizens who are well informed compared to those who are not. These findings are problematic from a normative perspective, since they indicate that differences in political information can result in unequal representation in the political system (see also Althaus, 1998; Kuklinski et al., 2000; Gilens, 2001). However, rather than relying on the degree to which individuals possess constrained belief systems, Delli Carpini and Keeter (1996) conceptualized knowledge as the awareness of key democratic values, using factual knowledge questions (see also Carpini and Keeter, 1993). Many studies focus on similar factual knowledge measures (e.g. Zaller, 1991; Gomez and Wilson, 2001). For example, Zaller (1992) measured political awareness using tests of neutral factual information about politics, since they "more directly than any of the alternative measures, capture what has actually gotten into peoples minds" (Zaller, 1992, 21). However, other research casts doubt on this assertion, both from methodological as well as theoretical perspectives.

From a methodological perspective, many studies raised issues related to the validity of factual knowledge questions as a measure of political sophistication. One problem are potential biases due to guessing (Mondak, 2000, 2001; Mondak and Davis, 2001; Miller and Orr, 2008). Responses to knowledge items that offer a "Don't Know" option are simultaneously influenced individual information levels as well as the propensity to guess. Other scholars further criticized open-ended factual knowledge questions due to problematic coding rules, which do not capture partial

knowledge (Krosnick et al., 2008; Gibson and Caldeira, 2009; DeBell, 2013).

Focusing solely on factual political knowledge has also been criticized on theoretical grounds. Lupia (2006) argued that the information asked for has no clear relevance to political participation. Instead, researchers should concentrate on heuristics that directly help citizens to make competent political decisions or focus only on knowledge relevant to a specific task (see also Lupia, 1994). Accordingly, there is no need for individuals to know all available facts, but only to possess the skills and resources to be able to *find* the information required in a specific context (Prior and Lupia, 2008). Furthermore, conventional items differ with regard to the dimension of political knowledge they measure (Barabas et al., 2014) and ignore important aspects such as visual cues (Prior, 2014).

Measuring Sophistication in Open-Ended Survey Responses

Rather than trying to develop a new item battery that addresses some of these issues, I propose an alternative approach. Instead of testing respondents on a specific set of predetermined facts, we can make inferences about political sophistication by analyzing how they discuss their attitudes and beliefs in their own words.

Let's return to our initial discussion of the theoretical concept of interest – political sophistication. As described above, Converse (1964) focused on the level of constraint in political beliefs rather than isolated pieces of factual information. Similarly, Tetlock (1983) used the term *integrative complexity* to describe the degree to which considerations related to an issue are interconnected. These studies do not conceptualize sophistication based on the content (or accuracy) of related considerations but rather on its *structure*. Luskin (1987) also defined political sophistication based on the structure of individual belief systems, arguing that belief systems can vary on three separate dimensions: (1) their *size* – i.e. the number of cognitions, (2) their *range* – i.e. the dispersion of cognition over categories, and (3) their *constraint* – i.e. the extent to which cognitions are interconnected in a meaningful way. Political sophistication, in turn, is seen

4

as the conjunction of these dimensions: "A person is politically sophisticated to the extent to which his or her [political belief system] is large, wide-ranging, and highly constrained." (Luskin, 1987, 860).

How would such a highly sophisticated person discuss his or her political views as compared to a less politically sophisticated individual? Consider for example a survey where respondents are asked to describe their attitudes toward specific policies or candidates running for office in their own words. In such a scenario, the structure of individual political belief systems (i.e., size, range, and constraint) should be reflected in their verbatim responses to a set of open-ended items. In the following, I discuss three different attributes of open-ended survey responses that should be indicative of individual political sophistication as described by Luskin (1987) and others.

First of all, sophisticated individuals should be able to elaborate more on their political attitudes. If people possess a large, wide-ranging, and constrained belief system, they should be able to recall a large number of considerations related to political actors or issues. Verbatim responses of sophisticates can therefore be expected to have a greater overall **length**, which is measured as the logged word count for each individual over all prompts:

$$\text{length}_i = \frac{\log\left(\sum_{j=1}^{J} n_{ij}\right)}{\max\left[\log\left(\sum_{j=1}^{J} n_{ij}\right)\right]}, \tag{1}$$

where n_{ij} is the number of words in the response of individual i in response to question j. J denotes the set of all open-ended probes. I use the logged count to normalize the distribution of responses and divide it by the maximal response length in the data such that the resulting measure ranges from 0 to 1.

However, sophisticated individuals should not only be able to talk about their attitudes at greater lengths. It is also important to consider the content of their respective answers. If a respondents holds more diverse cognitions towards political actors or policies, we should observe a wider range of topics addressed in their responses rather than a focus on single issues. As such, verbatim responses of political sophisticates should display a greater degree of **topic diversity**,

5

which is conceptualized as the relative mean absolute difference in topic proportions:[1]

$$\text{topic diversity}_i = 1 - \frac{\sum_{k_1=1}^{K} \sum_{k_2=1}^{K} |\theta_{ik_1} - \theta_{ik_2}|}{2 \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \theta_{ik_1}}, \tag{2}$$

where θ_{ik} denotes the predicted proportion of topic k in the collection of responses by individual i. The variable ranges from 0 (response focuses on single topic), to 1 (every topic has the same proportion). Mathematically, this conceptualization is equivalent to the Gini-coefficient, which measures the degree of inequality in income distributions (although the direction has been reversed such that a value of 1 implies a perfectly equal distribution).

Lastly, sophisticated individuals should hold opinions about each political actor or policy that they are asked to discuss. As such, shophisticates should be able to express their attitudes towards each open-ended probe in terms of both, approval or disapproval. Responses that reflect high levels of sophistication should therefore display a greater level of **opinion diversity**, which is measured as the relative mean absolute difference of relative response lengths for each open-ended response:

$$\text{opinion diversity}_i = 1 - \frac{\sum_{j_1=1}^{J} \sum_{j_2=1}^{J} |p_{ij_1} - p_{ij_2}|}{2 \sum_{j_1=1}^{J} \sum_{j_2=1}^{J} p_{ij_1}}, \tag{3}$$

where $p_{ij} = \frac{n_{ij}}{\sum_{j=1}^{J} n_{ij}}$ is the proportion of words in the response of individual i to question j relative to the overall size of the individuals' response. Again, the variable ranges from 0 (only one question was answered) to 1 (all questions were answered with the same word length per answer).

Together, the three measures form a composite metric of political sophistication by calculating their respective average for each respondent. Like each individual component, the resulting **text-**

---

[1]Individual topic proportions are extracted from a structural topic model estimated using the stm package in R (Roberts et al., 2014). The number of topics was selected using the algorithm of Lee and Mimno (2014) and the model was estimated via spectral initialization to address the issue of multi-modality (see Roberts, Stewart, and Tingley, 2014, for details). I use measures for age, education, party identification, as well as an interaction between education and party identification as covariates for topic prevalence. This variable selection is equivalent to the procedure model specification described in Roberts et al. (2014).

based sophistication score ranges from 0 to 1:

$$\text{text-based sophistication}_i = \tfrac{1}{3}(\text{length}_i + \text{topic diversity}_i + \text{opinion diversity}_i). \qquad (4)$$

Overall, a highly sophisticated individual can be expected to respond to a set of open-ended items by giving a more elaborate response that focuses on multiple considerations or topics and addresses his or her attitudes towards all relevant political actors or policies more or less equally.

Application 1: The Gender Gap in Political Sophistication

The following analyses are based on the 2012 American National Election Study (ANES), which consists of a representative survey of 5914 adults in the months before the 2012 US Presidential election. 2054 respondents participated in face-to-face interviews while the remaining 3860 filled out the survey online. For the purpose of the present analysis, I rely on the pooled dataset while controlling for differences in survey mode.

The text-based sophistication measure is based on open-ended questions in which respondents were asked in the pre-election wave of the survey 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. They were probed by the interviewer asking "anything else?" until the respondent answered "no". Overall, there are a total number of 8 open-ended responses where individuals described their beliefs and attitudes towards political actors. Verbatim responses were pre-processed using the Aspell spell-checking algorithm (www.aspell.net). Individuals who did not respond to all of the open-ended items (417 individuals), or who responded in Spanish (228 individuals), were excluded from the analysis.[2]

In the following, the text-based sophistication measure is compared to multiple alternative metrics of political knowledge. The most common way to measure political knowledge in surveys

---

[2]Please see Appendix A for descriptive information on the individual components of the text-based sophistication measure.

is to ask a set of factual questions about political institutions. The 2012 ANES includes such a basic item battery, inquiring for example about the number of times an individual can be elected President of the United States, or how the current U.S. federal budget deficit compares to the deficit in the 1990s. I combine individual responses on these items to a standard additive measure of **factual knowledge** about politics. Additionally, the in-person sample of the 2012 ANES includes **interviewer assessments** of each respondent's political sophistication.

Validating the Measure

Figure 1 provides a first validation of text-based sophistication by comparing it to the conventional knowledge metrics. The figure presents scatterplots between individual measures (lower triangular), univariate densities (diagonal), and correlation coefficients (upper triangular). The text-based sophistication measure is positively correlated with both conventional metrics while capturing some additional variation.

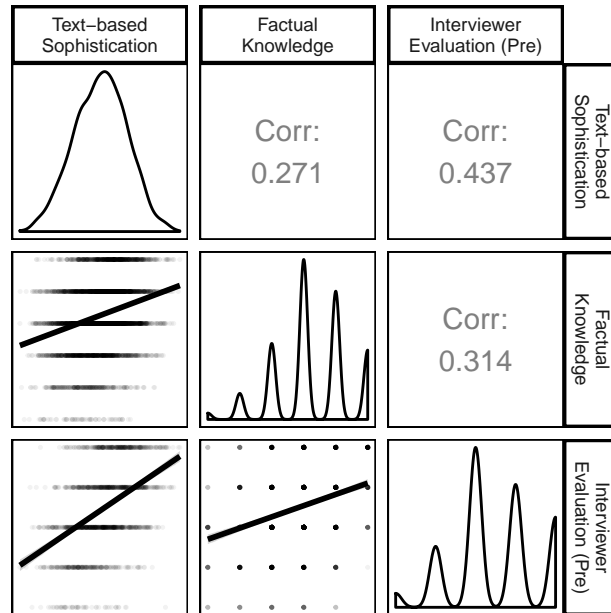Figure 1: Correlation matrix of conventional political knowledge metrics and the text-based sophistication measure. The plots on the diagonal display univariate densities for each variable. The panels in the lower triangular display the scatter plot of two measures as well as a linear fit. The upper triangular displays the correlation coefficient. All correlations reported are statistically significant with $p < .05$.

Interestingly, we observer a stronger correlation between text-based sophistication and interviewer evaluations than between factual knowledge and interviewer evaluations. The text-based measure therefore appears to capture characteristics that influence subjective assessments of sophistication. The interviewers certainly form their impressions throughout the entire survey, but it the complexity of a respondent's verbatim answers seems to be more influential than their performance on the factual knowledge questions.

Overall, while text-based sophistication and the alternative measures are clearly correlated, the relationship between each metric is far from perfect. To provide some intuition whether the variation in text-based sophistication is theoretically meaningful, I present an example of open-ended responses of two individuals who scored equally on the factual knowledge score (3 out of 5 correct responses), but varied highly in text-based sophistication. The results are presented in Table 1.

Each row in the table represents one of the open-ended responses (like/dislike for each candidate/party). Column A displays the responses of an individual who scored low on text-based sophistication and column B displays the responses of a high scoring individual. Cells are empty if a respondent refused to provide a response. Even though both individuals received are measured to be equal in their factual political knowledge, there are systematic differences in their response behavior that can be attributed to their political sophistication. Overall, respondent A provided a less elaborate response, only focused on two issues (health care and student loans), and did not report attitudes on multiple items. Compared to B, such a response pattern is suggestive of a less sophisticated political belief system. This initial validation suggests that the variation in the sophistication measure captures meaningful differences in response behavior that clearly overlaps with traditional knowledge metrics while displaying some unique variation.

Political sophistication is not only of theoretical interest as a dependent variable in political science research, but commonly used as a determinant of other outcomes related to attitudes and behavior. As a last validation of text-based sophistication, Figure 2 presents the effects of each sophistication measure on four dependent variables commonly related to political sophistication:

	A: Low Sophistication Response	B: High Sophistication Response
Obama (+)	The healthcare, keeping that and the financial aid, helping students.	I think he is honest, has good intentions.
Obama (-)		I don't feel he is up for the job, he doesn't really know how to get things accomplished from idea to actual reality.
Romney (+)		He comes across as an honest person and I feel that financially he would be better for the country.
Romney (-)	By taking financial aid away from students, taking family type planning, healthcare type of help away.	I am a moderate conservative and there are some things about anti-gay rights that I don't support.
Democrats (+)	Mostly the healthcare, mostly people do need healthcare and can't afford to pay insurance. Financial aid most people cant afford to go college. Main two things that I like is the help with education and to pay for insurance to go to doctor.	They do seem to be generally concerned with everyone, taking care of the country as a whole.
Democrats (-)		They fight too much among themselves and I disagree with wealth redistribution.
Republicans (+)		I agree with a lot of the conservative values and taking responsibility for one's own actions.
Republicans (-)		They argue too much among themselves and don't accomplish very much.

Table 1: Example of open-ended responses for low and high scores on the text-based sophistication measure with equal factual knowledge scores (3 out of 5 correct responses). Column A displays the verbatim responses of an individual who scored low on the text-based sophistication measure and column B displays the verbatim responses of an individual who scored high on the text-based sophistication measure. Each row represents one of the likes/dislikes items included in the analysis. Note that the responses in this table were slightly redacted for readability (spelling errors removed, etc.).

internal efficacy, external efficacy, non-conventional participation, and turnout. The results for the first three dependent variables are based on linear regressions while the effects on turnout are estimated using a logit model. Each model equation includes a single sophistication measure while controlling for gender, education, income, age, race, religiosity, and survey mode (face-to-face vs. online). Each plot displays the difference in the expected value of the respective dependent variable for maximum and minimum values of each sophistication measure, while holding all other variables at their means.

Overall, the sophistication metrics perform similarly as predictors of internal efficacy, external efficacy, non-conventional participation, and turnout. The effect of text-based sophistication
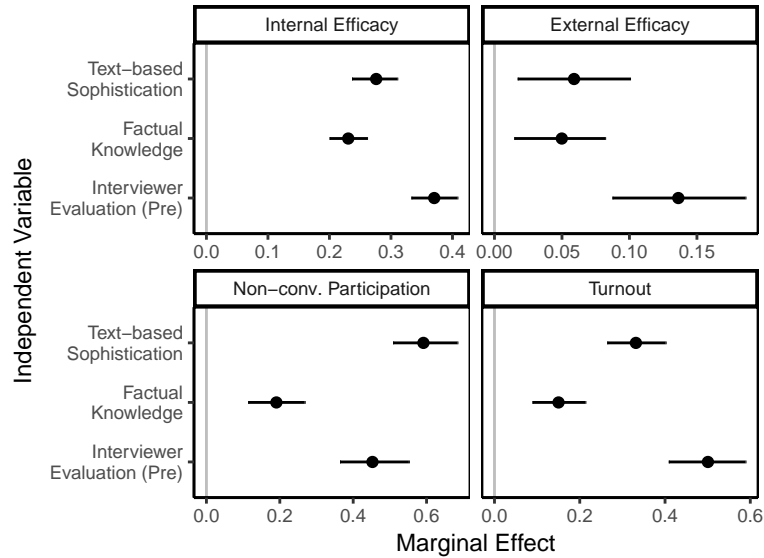
Figure 2: Effects of sophistication on internal efficacy, external efficacy, non-conventional partic-ipation, and turnout. For each dependent variable, the figure displays the difference in expected values between maximum and minimum levels of sophistication observed on each measure (in-cluding 95% confidence intervals). Model estimates are based on OLS (internal efficacy, external efficacy, non-conventional participation) or logistic regressions (turnout). Each sophistication measure is included in a single equation while controlling for gender, education, income, age, race, church attendance, and survey mode. Full model results ar presented in the appendix, Tables C.1 through C.4

on the participation measures is even stronger than the effect factual knowledge. This finding is especially noteworthy since the item batteries to measure factual political knowledge in the ANES are selected and validated based on their strong relationship with turnout and participation.

Assessing the Gender Gap

How do men and women compare on the different metrics of political sophistication? Figure 3 displays the average levels of text-based sophistication as well as the remaining metrics comparing both genders. While we observe a sizable gender gap for both conventional political knowledge measures, the difference is substantially smaller for text-based sophistication. Here, the gen-der gap is still statistically significant, but substantively inconsequential when compared to the remaining measures.

As described above, at least part of the gender gap can be attributed to real differences
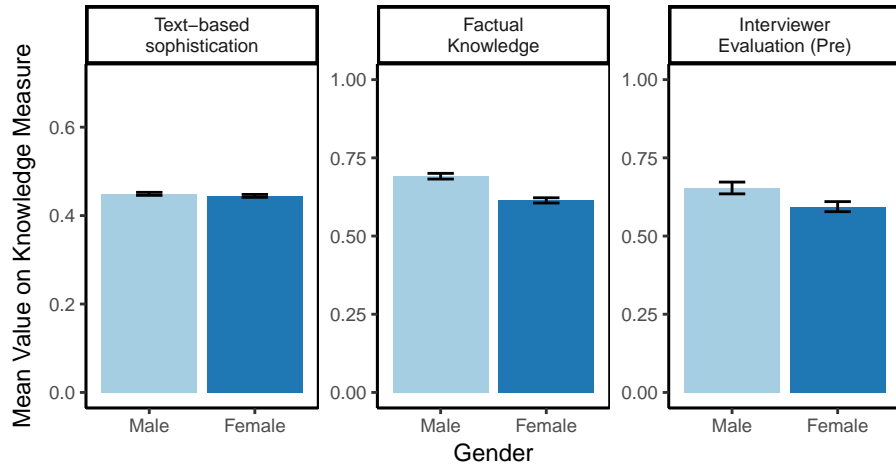
Figure 3: The gender gap in political sophistication. The figure displays mean levels of sophistication for each measure comparing men and women (including 95% confidence intervals). The y-axis is scaled to range up to the maximum value observed in the data for each sophistication metric. All gender differences are statistically significant with $p < .05$.

in resources relevant to political information (e.g., education). Accordingly, we need to control for common determinants of political knowledge across all available measures to provide a more comprehensive examination of potential gender differences. Previous studies consistently showed that political knowledge is positively related to high media exposure, frequent political discussions, education, and income. Furthermore, I include age, race, religiosity, and survey mode (face-to-face vs. online) as additional control variables. Figure 4 displays the coefficients of regression models with each knowledge/sophistication measure as the dependent variable.

After controlling for common determinants, text-based sophistication reveals no significant differences between men and women. On the other hand, we still observe the gender gap using all remaining political knowledge metrics considered here. As such, women might not score as highly on political quizzes (partly because they are less likely to guess rather express lack of knowledge), but they do not differ substantially in complexity and sophistication when they describe their political preferences.

The patterns for the remaining determinants are quite similar across different dependent variables. Knowledge and sophistication is significantly higher among respondents who are more exposed to political news media, discuss politics frequently, are more educated, and have higher
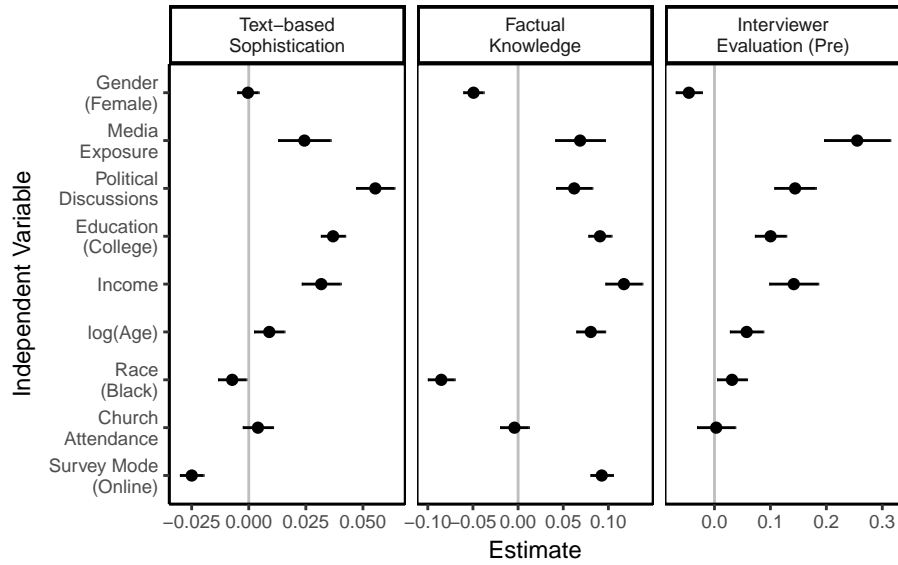
Figure 4: Common determinants of political sophistication. Estimates are OLS regression co-efficients with 95% confidence intervals. Dependent variables are the text-based sophistication measure as well as conventional metrics of political knowledge. Full model results are presented in the appendix, Table C.5

income. An interesting deviation, however, is the effect of survey mode. For factual knowledge questions, we observe that respondents in online surveys score significantly higher than individuals in face-to-face interviews. This difference could be explained by the fact that individuals are able to look up responses to factual knowledge questions while taking an online survey (see also Clifford and Jerit, 2016). For the text-based measure, on the other hand, we see that individuals appear to score lower on sophistication in online surveys. Respondents in online surveys therefore seem less willing to elaborate on their attitudes. Overall, the fact that the determinants of political sophistication are very consistent across models lends additional validity to the text-based measure.

As an additional step, we can examine whether women and men benefit equally from media exposure, political discussion, and education according to the distinct outcome measures. I extend the models from Figure 4 by including interactions between gender and media exposure, gender and political discussion, as well as gender and education. For simplicity, I only focus on the comparison of text-based sophistication with the conventional factual knowledge measure.

Figure 5 displays the expected sophistication levels for men and women on both metrics while increasing exposure, discussion, and education from their respective minimum to their maximum values (holding all other variables constant at their means).
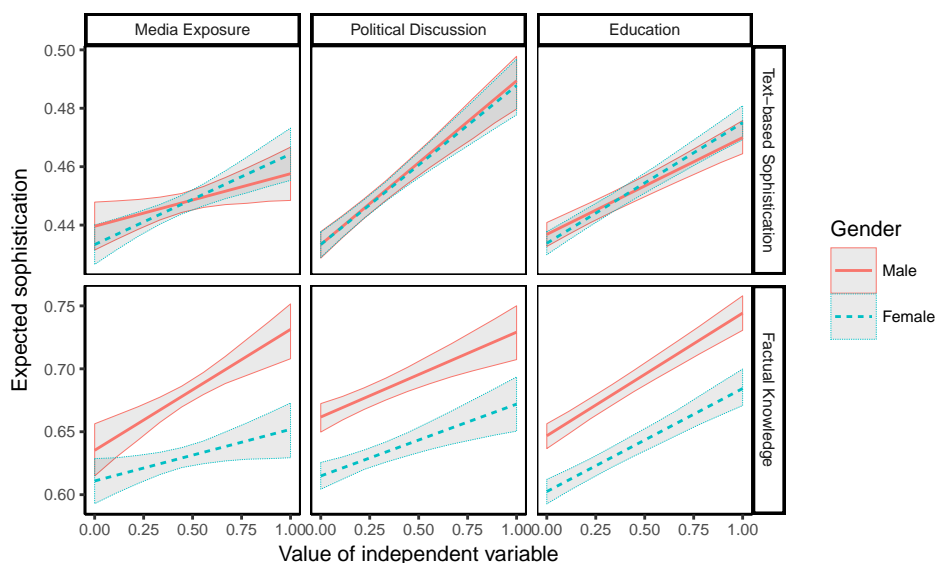

Figure 5: Closing the gender gap. The figure displays the expected values of text-based sophistication and factual knowledge for men and women depending on their respective levels of media exposure, political discussion, and education (including 95% confidence intervals). Estimates are based on OLS including interactions between gender and each independent variable. All models additionally control for education, age, race, church attendance, and survey mode. Full model results are presented in the appendix, Table C.6.

Media exposure, political discussions, and education increase political sophistication among men and women – irrespective of the specific measurement. However, there are important differences between the text-based measure and factual knowledge. For the conventional metric, we observe that the gender gap is not reduced with higher levels of media exposure, political discussion, or education. Indeed, media exposure even appears to increase gender differences in factual knowledge. This result seems inconsistent with recent research indicating that information exposure can mitigate the gender gap (e.g., Jerit et al., 2017). However, the growing difference in factual knowledge between men and women due to more frequent media exposure might be explained by the fact that the item battery does not capture gender-relevant (Dolan, 2011) or more practical (Stolle and Gidengil, 2010) political knowledge. For the text-based sophistication

14

measure, on the other hand, women benefit at least as much as men (if not more) from media exposure, political discussion, and education. Overall, solely focusing on factual knowledge may therefore lead to inaccurate conclusions regarding potential ways to mitigate gender differences in political sophistication.

Application 2: 2015 YouGov Survey

One of the advantages of measuring sophistication based on open-ended responses is that it can be applied in various settings that potentially focus on different types of political attitudes. Indeed, the logic of the text-based sophistication measure is not restricted to the type of open-ended items included in the 2012 ANES. From a theoretical perspective, the same arguments regarding the structure of individual belief systems holds when examining other types of open-ended responses, for example when respondents discuss their attitudes toward specific policy issues.

The following analysis will replicate and extend the main results of the previous section using a different survey employing an alternative set of open-ended responses. The data consist of a YouGov survey of 1000 US citizens that was conducted in December 2015. As part of this study, respondents were asked to report in a closed format whether they favor or oppose stricter gun laws. Subsequently, they are asked to respond to the following two questions in verbatim:

- 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.

- 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?

Similarly, the respondents report on their attitudes towards the Affordable Care Act in a closed format and are 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.

15

- 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?

Here, the text-based sophistication measure is computed based on the verbatim responses to the four preceding questions using the same procedures described in Study 1. Compared to the open-ended likes/dislikes items included in the 2012 ANES, the questions directly address considerations related to specific policy issues.[3]

Equivalent to the previous analysis, I begin by comparing text-based sophistication to a conventional political knowledge metric. The **factual knowledge** score in the YouGov data is based on a battery of eight items similar to the knowledge questions in the ANES. Additionally, the YouGov study included a task where respondents read a newspaper article about a fictional infectious disease and were subsequently asked to recall information provided in the article (e.g. regarding symptoms, modes of contraction etc.). I compute an additive index counting the pieces of information that were correctly recalled (**disease information**) as a measure of the ability to retrieve information from a news article on a non-partisan issue that is related to public health policies.

Validating the Measure

Figure 6 examines the distribution of each measure of political knowledge as well as their respective correlations in the YouGov data. Again, we observe that text-based sophistication and factual political knowledge are positively correlated, which indicates that the text-based sophistication measure overlaps with traditional knowledge metrics while capturing unique variation in individual response behavior. Interestingly, there appears to be a stronger relationship between text-based sophistication and disease information than between factual political knowledge and

---

[3]Please see Appendix B for descriptive information on the individual components of the text-based sophistication measure.
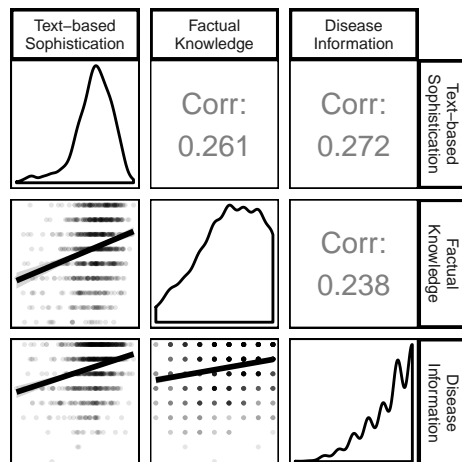
disease information.

Figure 6: YouGov data – Correlation matrix of text-based sophistication, a conventional political knowledge metric, and disease information retrieval. The plots on the diagonal display univariate densities for each variable. The panels in the lower triangular display the scatter plot of two measures as well as a linear fit. The upper triangular displays the correlation coefficient. All correlations reported are statistically significant with $p < .05$.

Recall that the disease information score can be interpreted as an indicator of the ability to retrieve specific information provided in a news article about a public health issue. It could be argued that text-based sophistication is more similar to the disease information score in that it captures the extent to which participants were able to recall political information that is relevant to their own attitudes. Conventional knowledge scores, on the other hand, inquire about specific that are not necessarily relevant to derive well-informed attitudes about political issues.

Assessing the Gender Gap

Next, I examine whether the results regarding the gender gap observed in Study 1 are replicated in the YouGov data. Figure 7 compares the alternative measures of political knowledge and sophistication for men and women. As before, we observe a significant gender gap in factual knowledge (middle panel), which disappears using the text-based measure (left panel). Furthermore, there is no apparent gender gap in disease information retrieval. If anything, women are able to recall slightly more accurate information about the infectious disease described in the news article. This

result is consistent with recent research by Jerit et al. (2017), which suggests that knowledge gaps between men and women disappear upon receiving sufficient information.
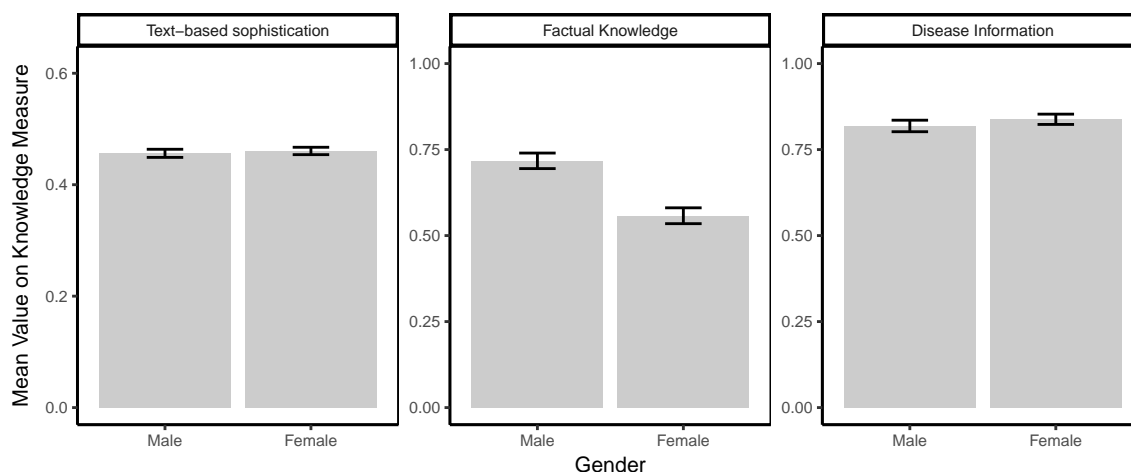
Figure 7: YouGov data – The gender gap in political sophistication. The figure displays mean levels of sophistication for each measure comparing men and women (including 95% confidence intervals). The y-axis is scaled to range up to the maximum value observed in the data for each sophistication metric. The gender difference in factual knowledge is statistically significant with $p < .05$.

A skeptical reader may again argue that the difference in factual knowledge between men and women could potentially be explained by systematic differences in resource-related factors. As such, we need to examine whether the gender gap persists even after controlling for covariates that might influence political knowledge. Similar to the analyses discussed in Study 1, Figure 8 presents the results of linear regressions predicting each measure as a function of gender, education, income, age, race, and religiosity.

As before, men do not perform better than women on text-based sophistication or disease information retrieval in a multivariate setting. However, the gender gap in factual political knowledge persists and is substantively as well as statistically significant. The remaining determinants of sophistication/knowledge are similar across measures (except for family income). Interestingly, the observed pattern of effects on text-based sophistication is strikingly similar to the effects on information retrieval about the fictional disease. This result reinforces the conclusion that text-based sophistication and disease information retrieval in the YouGov study share common
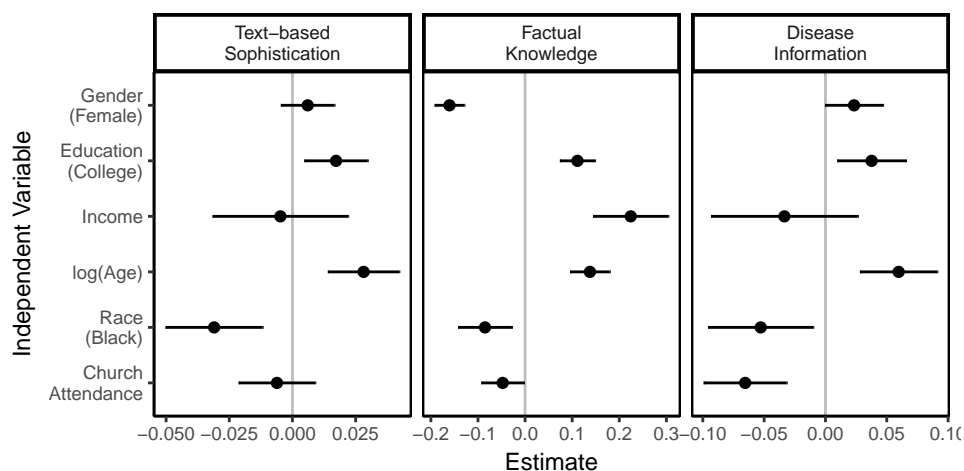
Figure 8: Common determinants of political sophistication – Replication using YouGov data. Estimates are OLS regression coefficients with 95% confidence intervals. Dependent variables are the text-based sophistication measure, factual political knowledge as well as knowledge about a fictional disease that was presented as part of the study. Full model results are presented in the appendix, Table C.7

characteristics in the sense that they capture how individuals recall considerations that are relevant to their own attitudes rather than inquiring about facts related to political institutions and elites that are extraneous to the issues at hand.

Conclusion

As Arthur Lupia (2015) described in his recent book "Uninformed", political scientists should worry less about pure levels of *information*, but rather focus on the necessary conditions for individuals to make *competent* decisions. Competence in the context of political decision-making and voting requires citizens to hold informed attitudes about their representatives. Factual knowledge about political institutions might be a useful proxy for competence in certain scenarios. However, it cannot address directly whether individuals are sufficiently opinionated about political actors they try to hold accountable. In comparison, the text-based sophistication measure proposed here is agnostic about the specific contents of individual beliefs, but directly captures the structure and complexity in attitude expression.

The findings presented in this paper show that conventional knowledge indices and the text-based measure share a substantial amount of variance. However, they are far from being identical and capture different aspects of sophistication. Most importantly, using the text-based measure, any evidence for the gender gap commonly reported using factual knowledge scales disappears. Women might know fewer facts about political institutions, but they do not differ substantively in the complexity of their expressed political beliefs.

References

Althaus, Scott L. 1998. "Information effects in collective preferences." *American Political Science Review* 92 (3): 545-558.

Barabas, Jason, Jennifer Jerit, William Pollock, and Carlisle Rainey. 2014. "The Question(s) of Political Knowledge." *American Political Science Review* 108 (04): 840–855.

Bartels, Larry M. 2005. "Homer gets a tax cut: Inequality and public policy in the American mind." *Perspectives on Politics* 3 (1): 15–31.

Bartels, Larry M. 2007. "Homer gets a warm hug: A note on ignorance and extenuation." *Perspectives on Politics* 5 (4): 785–790.

Carpini, Michael X Delli, and Scott Keeter. 1993. "Measuring political knowledge: Putting first things first." *American Journal of Political Science* 37 (4): 1179–1206.

Clifford, Scott, and Jennifer Jerit. 2016. "Cheating on Political Knowledge Questions in Online Surveys: An Assessment of the Problem and Solutions." *Public Opinion Quarterly*.

Converse, Philip E. 1964. "The Nature of Belief Systems in Mass Publics." In *Ideology and Discontent*, ed. David E. Apter. New York: Free Press.

DeBell, Matthew. 2013. "Harder than it looks: Coding political knowledge on the ANES." *Political Analysis* 21 (4): 393–406.

Delli Carpini, Michael X., and Scott Keeter. 1996. *What Americans Know about Politics and Why It Matters*. New Haven, London: Yale University Press.

Dolan, Kathleen. 2011. "Do women and men know different things? Measuring gender differences in political knowledge." *The Journal of Politics* 73 (01): 97–107.

Gibson, James L, and Gregory A Caldeira. 2009. "Knowing the Supreme Court? A reconsideration of public ignorance of the high court." *The Journal of Politics* 71 (02): 429–441.

Gilens, Martin. 2001. "Political ignorance and collective policy preferences." *American Political Science Association* 95 (02): 379–396.

Gomez, B.T., and J.M. Wilson. 2001. "Political sophistication and economic voting in the American electorate: A theory of heterogeneous attribution." *American Journal of Political Science* 45 (4): 899–914.

Jerit, Jennifer et al. 2017. "Revisiting the Gender Gap in Political Knowledge." *Political Behavior* forthcoming.

Krosnick, Jon A, Arthur Lupia, Matthew DeBell, and Darrell Donakowski. 2008. Problems with ANES questions measuring political knowledge. Technical report Ann Arbor, MI: American National Election Studies Report.

Kuklinski, James H, Paul J Quirk, Jennifer Jerit, David Schwieder, and Robert F Rich. 2000. "Misinformation and the currency of democratic citizenship." *Journal of Politics* 62 (3): 790–816.

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.

Lupia, Arthur. 1994. "Shortcuts Versus Encyclopedias: Information and Voting Behavior in California Insurance Reform Elections." *American Political Science Review* 88 (1): 63-76.

Lupia, Arthur. 2006. "How elitism undermines the study of voter competence." *Critical Review* 18 (1-3): 217–232.

Lupia, Arthur. 2015. *Uninformed: Why people seem to know so little about politics and what we can do about it*. Oxford University Press.

Lupia, Arthur, Adam Seth Levine, Jesse O Menning, and Gisela Sin. 2007. "Were Bush tax cut supporters simply ignorant? A second look at conservatives and liberals in Homer gets a tax cut." *Perspectives on Politics* 5 (4): 773–784.

Luskin, Robert C. 1987. "Measuring political sophistication." *American Journal of Political Science* 31 (4): 856–899.

Miller, Melissa K, and Shannon K Orr. 2008. "Experimenting with a Third Way in political knowledge estimation." *Public Opinion Quarterly* 72 (4): 768–780.

Mondak, Jeffery J. 2000. "Reconsidering the measurement of political knowledge." *Political Analysis* 8 (1): 57–82.

Mondak, Jeffery J, and Belinda Creel Davis. 2001. "Asked and answered: Knowledge levels when we will not take 'don't know' for an answer." *Political Behavior* 23 (3): 199–224.

Mondak, Jeffery J, and Mary R Anderson. 2004. "The knowledge gap: A reexamination of gender-based differences in political knowledge." *Journal of Politics* 66 (2): 492–512.

Mondak, Jeffrey J. 2001. "Developing valid knowledge scales." *American Journal of Political Science* 45 (1): 224–238.

Pietryka, Matthew T, and Randall C MacIntosh. 2013. "An Analysis of ANES Items and Their Use in the Construction of Political Knowledge Scales." *Political Analysis* 21 (4): 407–429.

Prior, Markus. 2014. "Visual political knowledge: A different road to competence?" *Journal of Politics* 76 (1): 41–57.

Prior, Markus, and Arthur Lupia. 2008. "Money, time, and political knowledge: Distinguishing quick recall and political learning skills." *American Journal of Political Science* 52 (1): 169–183.

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.

Stolle, Dietlind, and Elisabeth Gidengil. 2010. "What do women really know? A gendered analysis of varieties of political knowledge." *Perspectives on Politics* 8 (01): 93–109.

Sturgis, Patrick, Nick Allum, and Patten Smith. 2008. "An experiment on the measurement of political knowledge in surveys." *Public Opinion Quarterly* 72 (1): 90–102.

Tetlock, Philip E. 1983. "Cognitive style and political ideology." *Journal of Personality and Social Psychology* 45 (1): 118.

Verba, Sidney, Nancy Burns, and Kay Lehman Schlozman. 1997. "Knowing and caring about politics: Gender and political engagement." *The Journal of Politics* 59 (04): 1051–1072.

Wolak, Jennifer, and Michael McDevitt. 2011. "The roots of the gender gap in political knowledge in adolescence." *Political Behavior* 33 (3): 505–533.

Zaller, John. 1991. "Information, values, and opinion." *American Political Science Review* 85 (04): 1215–1237.

Zaller, John. 1992. *The nature and origins of mass opinion*. Cambridge University Press.

Appendix A: Open-ended Responses in ANES
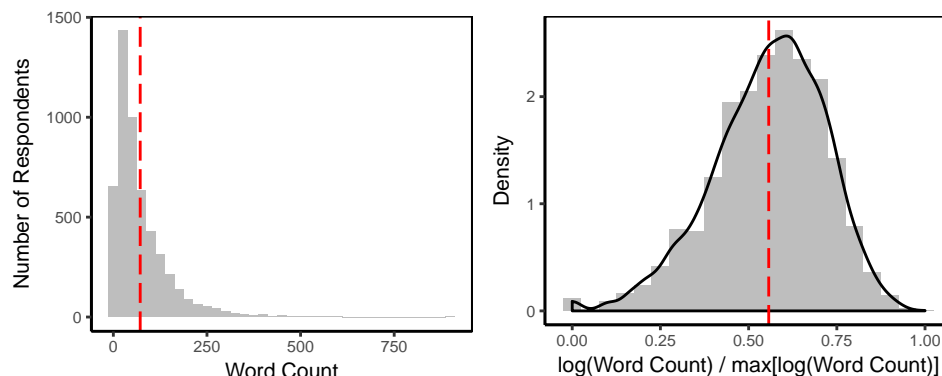
Figure A.1: Histogram of total word count in the collection of open-ended responses for each individual (left panel) and distribution of logged word count used in the text-based sophistication measure (right panel, re-scaled to range from 0 to 1). Dashed red lines indicate mean values. Most respondents provide brief statements when they describe their attitudes towards political parties and candidates. The mean response length to all 8 questions is about 75 words, so an average response to a single question consisted of less than 10 words, omitting respondents who did not provide any information.
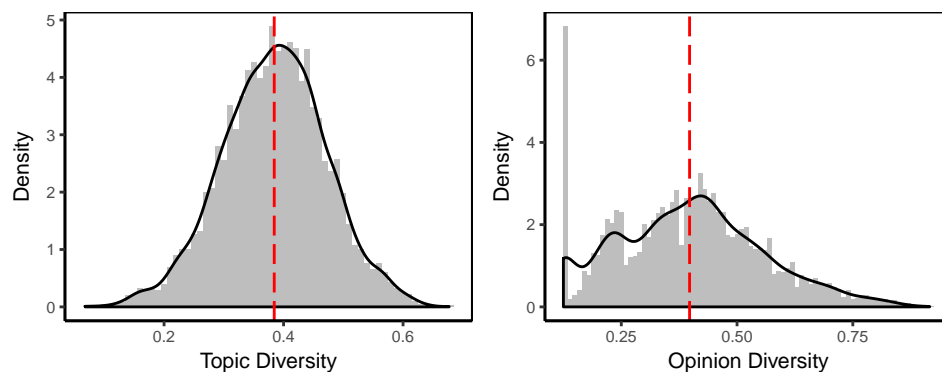
Figure A.2: Histogram and density of the topic diversity (left panel) and opinion diversity (right panel) measure. Dashed red lines indicate mean values. The spike at 0 for opinion diversity are due to the fact that a large proportion of respondents only answered a single open-ended question.

Appendix B: Open-ended Responses in YouGov Data
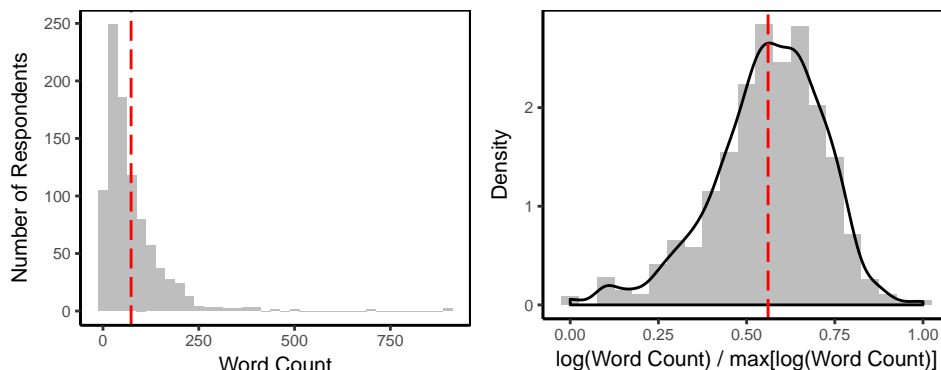
Figure B.1: Histogram of total word count in the collection of open-ended responses for each individual (left panel) and distribution of logged word count used in the text-based sophistication measure (right panel, re-scaled to range from 0 to 1). Dashed red lines indicate mean values. Most respondents provide brief statements when they describe their attitudes towards political parties and candidates. The mean response length to all 8 questions is about 73 words, so an average response to a single question consisted of approximately 18 words, omitting respondents who did not provide any information.
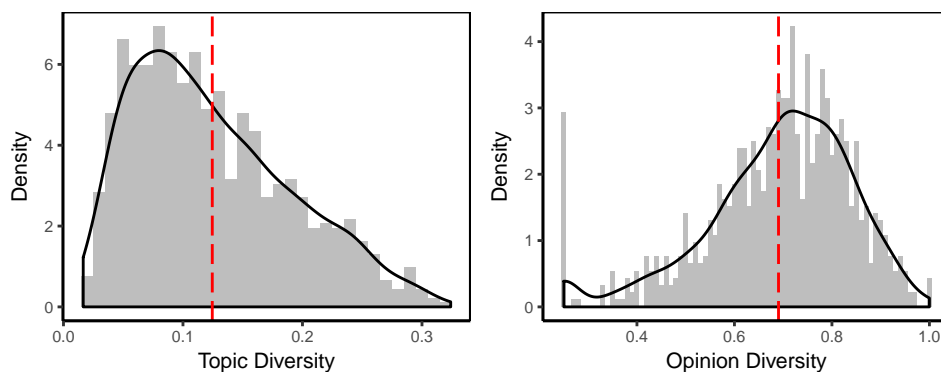
Figure B.2: Histogram and density of the topic diversity (left panel) and opinion diversity (right panel) measure. Dashed red lines indicate mean values. The spike at 0 for opinion diversity are due to the fact that a large proportion of respondents only answered a single open-ended question.

Appendix C: Tables of Model Estimates

Table C.1: Effects of sophistication – OLS models predicting internal efficacy based on different sophistication measures. Positive coefficients indicate higher self-reported internal efficacy. Standard errors in parentheses. Estimates are used for Figure 2 in the main text.

	Dependent variable:					
	Iternal Efficacy					
	(1)	(2)	(3)	(4)	(5)	(6)
Text-based	0.526***					
	(0.036)					
Factual		0.231***				
		(0.015)				
Office			0.244***			
			(0.012)			
Majorities				0.136***		
				(0.008)		
Eval. (Pre)					0.370***	
					(0.020)	
Eval. (Post)						0.286***
						(0.021)
Sex (Female)	−0.061***	−0.049***	−0.051***	−0.052***	−0.037***	−0.039***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.010)	(0.010)
Education (College)	0.048***	0.048***	0.034***	0.054***	0.036**	0.044***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.012)	(0.013)
Income	0.042***	0.033**	0.019	0.045***	−0.007	0.006
	(0.011)	(0.011)	(0.011)	(0.011)	(0.018)	(0.020)
log(Age)	0.017*	0.005	0.006	0.006	−0.020	−0.010
	(0.008)	(0.008)	(0.008)	(0.008)	(0.012)	(0.013)
Race (Black)	0.044***	0.059***	0.048***	0.044***	0.046***	0.055***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.011)	(0.012)
Church Attendance	0.005	0.009	0.013	0.005	−0.002	−0.007
	(0.008)	(0.008)	(0.008)	(0.009)	(0.014)	(0.015)
Survey Mode (Online)	0.042***	0.006	−0.011	−0.002		
	(0.007)	(0.007)	(0.007)	(0.007)		
Constant	0.188***	0.373***	0.449***	0.445***	0.391***	0.416***
	(0.034)	(0.031)	(0.031)	(0.032)	(0.045)	(0.049)
Observations	5,004	5,004	4,706	4,706	1,662	1,580
R^2	0.114	0.117	0.158	0.128	0.232	0.169

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

Table C.2: Effects of sophistication – OLS models predicting external efficacy based on different sophistication measures. Positive coefficients indicate higher self-reported external efficacy. Standard errors in parentheses. Estimates are used for Figure 2 in the main text.

	Dependent variable:					
	External Efficacy					
Text-based	0.110** (0.041)					
Factual		0.049** (0.017)				
Office			0.084*** (0.013)			
Majorities				0.044*** (0.010)		
Eval. (Pre)					0.136*** (0.026)	
Eval. (Post)						0.149*** (0.027)
Sex (Female)	0.014* (0.007)	0.016* (0.007)	0.015* (0.007)	0.015* (0.007)	0.023 (0.013)	0.023 (0.014)
Education (College)	0.041*** (0.008)	0.041*** (0.008)	0.033*** (0.008)	0.040*** (0.008)	0.064*** (0.016)	0.057*** (0.016)
Income	0.023 (0.012)	0.021 (0.012)	0.012 (0.013)	0.021 (0.013)	0.006 (0.025)	0.002 (0.025)
log(Age)	−0.008 (0.009)	−0.011 (0.009)	−0.013 (0.010)	−0.012 (0.010)	−0.051** (0.016)	−0.048** (0.017)
Race (Black)	0.077*** (0.009)	0.080*** (0.009)	0.078*** (0.009)	0.076*** (0.009)	0.058*** (0.015)	0.058*** (0.015)
Church Attendance	0.048*** (0.010)	0.049*** (0.010)	0.053*** (0.010)	0.050*** (0.010)	0.045* (0.019)	0.042* (0.020)
Survey Mode (Online)	−0.036*** (0.008)	−0.043*** (0.008)	−0.055*** (0.008)	−0.051*** (0.008)		
Constant	0.369*** (0.038)	0.407*** (0.035)	0.434*** (0.036)	0.432*** (0.037)	0.462*** (0.061)	0.455*** (0.063)
Observations	4,993	4,993	4,694	4,694	1,653	1,569
R^2	0.040	0.040	0.048	0.044	0.053	0.053

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

Table C.3: Effects of sophistication – OLS models predicting non-conventional particpation (protest, signing petitions, etc.) based on different sophistication measures. Positive coefficients indicate higher levels of participation. Standard errors in parentheses. Estimates are used for Figure 2 in the main text.

	Dependent variable:					
	Non-conventional Participation					
Text-based	1.124***					
	(0.085)					
Factual		0.193***				
		(0.037)				
Office			0.322***			
			(0.028)			
Majorities				0.151***		
				(0.020)		
Eval. (Pre)					0.453***	
					(0.050)	
Eval. (Post)						0.451***
						(0.050)
Sex (Female)	0.010	0.017	0.025	0.021	0.033	0.036
	(0.014)	(0.015)	(0.014)	(0.015)	(0.025)	(0.025)
Education (College)	0.082***	0.110***	0.082***	0.111***	0.103***	0.096**
	(0.017)	(0.017)	(0.017)	(0.016)	(0.030)	(0.030)
Income	0.063*	0.082**	0.048	0.087***	0.036	0.040
	(0.026)	(0.027)	(0.026)	(0.026)	(0.046)	(0.046)
log(Age)	0.036	0.038	0.026	0.031	−0.057	−0.054
	(0.019)	(0.020)	(0.019)	(0.020)	(0.031)	(0.031)
Race (Black)	0.040*	0.048*	0.042*	0.037	−0.020	−0.012
	(0.019)	(0.019)	(0.019)	(0.019)	(0.028)	(0.028)
Church Attendance	0.025	0.032	0.038	0.028	0.046	0.041
	(0.020)	(0.021)	(0.020)	(0.020)	(0.036)	(0.036)
Survey Mode (Online)	0.105***	0.055***	0.024	0.041*		
	(0.016)	(0.016)	(0.016)	(0.016)		
Constant	−0.415***	0.008	0.108	0.092	0.270*	0.284*
	(0.081)	(0.075)	(0.075)	(0.075)	(0.115)	(0.115)
Observations	4,701	4,701	4,701	4,701	1,573	1,573
R^2	0.071	0.042	0.063	0.048	0.083	0.082

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

Table C.4: Effects of sophistication – Logit models predicting turnout based on different sophistication measures. Positive coefficients indicate higher probabilities to participate in the election. Standard errors in parentheses. Estimates are used for Figure 2 in the main text.

	Dependent variable:					
	Turnout					
Text-based	4.960***					
	(0.497)					
Factual		1.087***				
		(0.201)				
Office			1.982***			
			(0.183)			
Majorities				1.089***		
				(0.110)		
Eval. (Pre)					2.795***	
					(0.282)	
Eval. (Post)						2.726***
						(0.285)
Sex (Female)	0.062	0.072	0.135	0.135	0.172	0.190
	(0.083)	(0.083)	(0.084)	(0.084)	(0.134)	(0.135)
Education (College)	0.510***	0.604***	0.438***	0.585***	0.494**	0.438*
	(0.108)	(0.107)	(0.109)	(0.107)	(0.185)	(0.184)
Income	1.075***	1.116***	0.949***	1.164***	1.107***	1.168***
	(0.152)	(0.152)	(0.155)	(0.152)	(0.259)	(0.257)
log(Age)	1.063***	1.035***	0.974***	0.974***	0.726***	0.703***
	(0.103)	(0.103)	(0.104)	(0.104)	(0.160)	(0.160)
Race (Black)	0.811***	0.864***	0.823***	0.804***	0.943***	0.979***
	(0.119)	(0.119)	(0.118)	(0.118)	(0.165)	(0.165)
Church Attendance	0.709***	0.733***	0.761***	0.714***	0.813***	0.757***
	(0.127)	(0.126)	(0.127)	(0.127)	(0.205)	(0.205)
Survey Mode (Online)	0.645***	0.388***	0.218*	0.236**		
	(0.088)	(0.087)	(0.089)	(0.090)		
Constant	−6.639***	−4.709***	−4.214***	−4.183***	−4.265***	−3.990***
	(0.450)	(0.391)	(0.395)	(0.395)	(0.604)	(0.598)
Observations	4,724	4,724	4,706	4,706	1,582	1,580
Akaike Inf. Crit.	3,811.953	3,885.718	3,777.952	3,809.053	1,470.846	1,477.271

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

Table C.5: Determinants of political knowledge – OLS models predicting different political sophistication measures. Positive coefficients indicate higher sophistication. Standard errors in parentheses. Estimates are used for Figure 4 in the main text.

	Dependent Variable: Political Knowledge Measure					
	Text-based	Factual	Office	Majorities	Eval. (Pre)	Eval. (Post)
Gender (Female)	−0.0004	−0.049***	−0.045***	−0.082***	−0.046***	−0.052***
	(0.002)	(0.006)	(0.007)	(0.011)	(0.012)	(0.012)
Media Exposure	0.024***	0.069***	0.159***	0.193***	0.255***	0.192***
	(0.006)	(0.014)	(0.018)	(0.026)	(0.030)	(0.030)
Political Discussions	0.055***	0.062***	0.195***	0.200***	0.144***	0.174***
	(0.004)	(0.010)	(0.013)	(0.019)	(0.019)	(0.019)
Education (College)	0.037***	0.091***	0.129***	0.099***	0.100***	0.118***
	(0.003)	(0.007)	(0.008)	(0.012)	(0.014)	(0.014)
Income	0.032***	0.117***	0.147***	0.091***	0.142***	0.135***
	(0.004)	(0.010)	(0.013)	(0.019)	(0.022)	(0.022)
log(Age)	0.009**	0.081***	0.048***	0.122***	0.057***	0.062***
	(0.003)	(0.008)	(0.010)	(0.015)	(0.015)	(0.015)
Race (Black)	−0.007*	−0.085***	−0.036***	−0.034*	0.031*	0.020
	(0.003)	(0.008)	(0.010)	(0.014)	(0.014)	(0.014)
Church Attendance	0.004	−0.004	−0.028**	0.014	0.003	0.014
	(0.003)	(0.008)	(0.010)	(0.015)	(0.017)	(0.017)
Survey Mode (Online)	−0.025***	0.093***	0.163***	0.214***		
	(0.003)	(0.006)	(0.008)	(0.012)		
Constant	0.402***	0.101***	−0.198***	−0.376***	0.199***	0.137*
	(0.013)	(0.030)	(0.038)	(0.055)	(0.056)	(0.055)
Observations	4,698	4,698	4,698	4,698	1,575	1,575
R^2	0.152	0.282	0.322	0.225	0.242	0.251

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

Table C.6: Closing the Gender Gap – OLS models predicting different political sophistication measures. Positive coefficients indicate higher sophistication. Standard errors in parentheses. Estimates are used for Figure 5 in the main text.

	Dependent Variable: Political Knowledge Measure	
	Text-based	Factual
Gender (Female)	−0.008	−0.016
	(0.005)	(0.013)
Media Exposure	0.018*	0.097***
	(0.008)	(0.020)
Political Discussions	0.056***	0.068***
	(0.006)	(0.014)
Education (College)	0.033***	0.098***
	(0.004)	(0.009)
Income	0.032***	0.117***
	(0.004)	(0.010)
log(Age)	0.009**	0.080***
	(0.003)	(0.008)
Race (Black)	−0.007*	−0.085***
	(0.003)	(0.008)
Church Attendance	0.004	−0.004
	(0.003)	(0.008)
Survey Mode (Online)	−0.025***	0.093***
	(0.003)	(0.006)
Female * Media	0.013	−0.056*
	(0.011)	(0.026)
Female * Discussions	−0.001	−0.011
	(0.008)	(0.020)
Female * Education	0.008	−0.016
	(0.005)	(0.012)
Constant	0.405***	0.087**
	(0.013)	(0.031)
Observations	4,698	4,698
R^2	0.153	0.284
Note:		*p<0.05; **p<0.01; ***p<0.001

Table C.7: Determinants of political knowledge (YouGov data) – OLS models predicting political sophistication and disease information retrieval. Positive coefficients indicate higher sophistication. Standard errors in parentheses. Estimates are used for Figure 8 in the main text.

	Dependent Variable: Political Knowledge Measure		
	Text-based	Factual	Disease
Gender (Female)	0.006	−0.160***	0.023
	(0.005)	(0.016)	(0.012)
Education (College)	0.017**	0.111***	0.038**
	(0.006)	(0.019)	(0.014)
Income	−0.005	0.224***	−0.033
	(0.014)	(0.041)	(0.031)
log(Age)	0.028***	0.138***	0.060***
	(0.007)	(0.021)	(0.016)
Race (Black)	−0.031**	−0.085**	−0.053*
	(0.010)	(0.029)	(0.022)
Church Attendance	−0.006	−0.048*	−0.065***
	(0.008)	(0.023)	(0.017)
Constant	0.350***	0.116	0.618***
	(0.028)	(0.083)	(0.062)
Observations	795	795	795
R^2	0.046	0.272	0.054
Note:			*p<0.05; **p<0.01; ***p<0.001