

Women Also Know Stuff

Challenging the Gender Gap in Political Sophistication*

Patrick W. Kraft[†]

Abstract

This paper proposes a simple but powerful framework to measure political sophistication based on open-ended survey responses. *Discursive sophistication* uses automated text analysis methods to capture the complexity of individual attitude expression. I validate the approach by comparing it to conventional political knowledge metrics using different batteries of open-ended items across five surveys spanning four languages (total $N \simeq 35,000$). The new measure casts doubt on the oft-cited gender gap in political knowledge: Women might know fewer facts about institutions and elites, but they do not differ substantively in the sophistication of their expressed political attitudes.

Keywords: political sophistication, gender gap, open-ended responses, text analysis

*This title is inspired by <https://womentalkpolitics.com/>, an organization that promotes the work of women in political science by providing a public database of relevant women experts for journalists, scholars, and others. Previous versions of this manuscript have been presented at Polmeth, MPSA, EPSA, ISPP, APSA, the Toronto Political Behavior Meeting, and Zurich Text as Data. I thank Alexa Bankert, Jason Barabas, Scott Clifford, Andy Delton, Milena Djourelova, Kathy Dolan, Peter DeScioli, Stanley Feldman, Roy Gava, Fabrizio Gilardi, Jennifer Jerit, Yanna Krupnikov, Emmy Lindstam, Hannah Nam, Michael Persess, Daniel Rubenson, Rüdiger Schmitt-Beck, and Arthur Spirling for helpful comments on previous versions of this manuscript. Special thanks to Céline Colombo, Scott Clifford, and Jennifer Jerit for sharing their data. The manuscript and code are available on GitHub: <https://github.com/pwkraft/knowledge>.

[†]Assistant Professor, University of Wisconsin-Milwaukee, kraftp@uwm.edu.

Political sophistication is one of the most fundamental concepts in the study of political attitudes and behavior. While scholars routinely emphasize people's alarmingly low levels of sophistication (Converse, 1964; Delli Carpini and Keeter, 1996), fundamental issues regarding its measurement continue to plague the discipline (Mondak, 2001; Sturgis, Allum, and Smith, 2008; Bullock and Rader, 2021). Scholars usually rely on survey questions that assess people's ability to recall basic facts about political institutions and officeholders as a proxy for sophistication (Carpini and Keeter, 1993; Barabas et al., 2014). In principle, these factual knowledge questions should cover information that is necessary for citizens to make competent decisions in a given context, but determining such a set of items proves to be extremely difficult (Lupia, 2006). Even within a given policy area, people may disagree about which facts are crucial for political competence due to inherent value differences (Lupia, 2015). Furthermore, different sets of knowledge questions vary in difficulty across subgroups of the population (i.e., differential item functioning), which can introduce systematic measurement error (Petryka and MacIntosh, 2013).

One manifestation of such systematic measurement error is the oft-cited gender gap in political sophistication. On the basis of conventional factual knowledge scores, women frequently appear to be less informed about politics than men (Verba, Burns, and Schlozman, 1997; Wolak and McDevitt, 2011; Fraile, 2014a). To a certain extent, these findings reflect genuine differences in political interest between men and women due to gendered socialization (Bos et al., 2021). However, at least part of the observed gender gap can be attributed to how we measure political knowledge. For instance, men tend to have a higher propensity to guess when answering recall questions, thus overestimating their knowledge levels (Mondak and Anderson, 2004; Fortin-Rittberger, 2020). Other research found that gender differences are attenuated when focusing on gender-relevant political knowledge (Dolan, 2011), by providing policy-specific information (Jerit and Barabas, 2017), or in contexts with higher women representation (Pereira, 2019).

In this paper, I re-examine the gender gap by proposing a measure of *discursive so-*

phistication that is based on how people discuss their political preferences in open-ended survey responses. Specifically, I develop a framework to assess whether beliefs and attitudes are expressed in a more elaborate manner—a question that is not directly discernible from off-the-shelf factual knowledge items. Measuring sophistication based on how people talk about politics provides two major advantages compared to off-the-shelf factual knowledge items: (1) it captures the extent to which a respondent’s political beliefs are based on elaborate reasoning, and (2) it can easily pinpoint competence in specific areas by incorporating targeted open-ended items. The resulting measurement is therefore conceptually closer to the underlying latent trait of interest: the degree of structure and constraint in political belief systems (Tetlock, 1983; Luskin, 1987). Furthermore, it can be easily applied by researchers interested in analyzing open-ended responses with an easy-to-use software package available in the statistical programming environment R.¹

I validate the measure across multiple representative surveys in the United States and Europe encompassing four languages (total $N \simeq 35,000$) by comparing it to conventional factual knowledge scores as predictors of various indicators of civic competence and engagement. While discursive sophistication shares a considerable amount of variance with traditional metrics, they are far from equivalent. Indeed, discursive sophistication is a stronger predictor of turnout and other forms of political engagement. Contrary to previous research, however, I find no evidence for a gender gap in discursive 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. Furthermore, I present suggestive evidence that this divergence can be explained by the fact that open-ended responses allow women to focus on different issues than men. In sum, the results suggest that exploring open-ended responses provides new opportunities to examine political sophistication across time and contexts.

¹Package under development, release on CRAN expected mid 2022.

Political Sophistication and Factual Knowledge

Public opinion researchers routinely incorporate political sophistication in their empirical analysis in one way or another—either directly as an outcome variable of interest, as a major explanatory factor, or as an important confounder to control for. In order to measure the underlying latent trait, scholars commonly rely on short batteries of standard recall questions on basic facts about the political system.² One canonical article proposing such a battery—[Carpini and Keeter \(1993\)](#)—has been cited more than 1000 times since its publication. Figure 1 shows the yearly citation count over time; the trend illustrates how political knowledge remains a concept of intense scholarly interest—and that it is frequently measured using standard off-the-shelf recall questions.

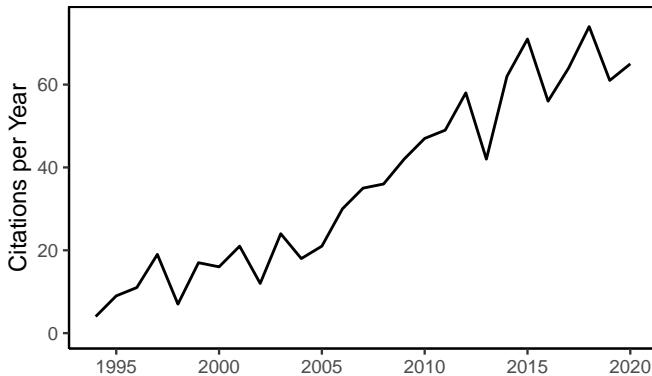


Figure 1: Yearly citation count of [Carpini and Keeter \(1993\)](#) based on Google Scholar.

The ubiquity of basic recall questions in public opinion research is accompanied by the frequent finding that many people know too little about politics ([Delli Carpini and Keeter, 1996](#); [Barabas et al., 2014](#)) and the discrepancies in information levels can result in unequal representation in the political system ([Althaus, 1998](#); [Kuklinski et al., 2000](#); [Gilens, 2001](#)). The underlying reason why scholars focus on people’s ability to recall factual information about politics is that these items “more directly than any of the alternative measures, capture what has actually gotten into peoples minds” ([Zaller 1992](#), 21; see also [Zaller](#)

²For example, the American National Election Study routinely asks questions such as, “Do you happen to know which party currently has the most members in the U.S. Senate?”

1991; Gomez and Wilson 2001). However, there is some reason to doubt this assertion, both from theoretical as well as methodological perspectives.

First, the discipline's focus solely on factual political knowledge has been criticized on theoretical grounds. Most importantly, recalling facts about political institutions has little relevance for citizen competence (Lupia, 2006; Cramer and Toff, 2017). Given that there is usually no consensus about what information is necessary in the first place, Druckman (2014) proposes abandoning recall questions as measures of "quality opinion." Instead, the author advocates "*less* focus on the *content/substance* of opinions [...] and *more* on the *process* and specifically the *motivation* that underlies the formation of those opinions" (2014, 478, emphasis in the original). The key distinction should therefore be how citizens approach a political issue and whether they are motivated to engage in elaborate reasoning to arrive at their particular decision. In addition, 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).

Second, there are several methodological issues that cast doubt on the validity of factual knowledge scores as a measure of political sophistication. One problem frequently discussed in the literature centers around the question whether or not to offer "Don't Know" options in multiple choice recall questions (Mondak, 2000; Mondak and Davis, 2001; Miller and Orr, 2008). Including such an option can lead to biased estimates of information levels because they are confounded by people's differential propensity to guess instead of admitting not to know the correct answer (but see Luskin and Bullock, 2011). Other scholars 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). The increasing reliance on online surveys instead of phone or face-to-face surveys creates additional concerns due to people's (differential) tendency to cheat by looking up answers for recall questions (Clifford and Jerit, 2016; Höhne et al., 2021; Style and Jerit, 2021). Lastly, factual knowledge scores have been shown to suffer from differential item functioning, since individual recall questions have varying measurement properties across the population (Pietryka and MacIntosh, 2013). Item batteries that are easier to answer for certain groups can therefore exacerbate observed differences in political knowledge—for example between racial groups (Abrajano, 2014).

Please Mind the Gender Gap

Survey researchers not only find that people are not sufficiently informed as a whole, they also attest that women are systematically less knowledgeable than men. For instance, Verba, Burns, and Schlozman (1997) report that women score lower on political information, interest, and efficacy, which decreases their respective levels of political participation. Since gender differences in political information and interest can only partly be explained by resource-related factors such as individual levels of education, the authors diagnose a “genuine difference in the taste for politics” between women and men, which they suspect is driven largely by socialization (see also Wolak and McDevitt, 2011). Indeed, Dow (2009, 117) describes the systematic gender differences in knowledge as “one of the most robust findings in the study of political behavior.” While differences between women and men in political *interest* can certainly be attributed to gendered political socialization (Bos et al., 2021; Wolak, 2020), at least part of the disparities in *knowledge* may simply be an artifact of the measurement approach.

The discussion revolving around the apparent gender gap is therefore closely intertwined with the methodological debate about measuring political knowledge. For in-

stance, [Mondak and Anderson \(2004\)](#) suggest that women are more likely to report that they do not know the answer to a recall question whereas men are more inclined to guess. Correcting for the systematic differences in the propensity to guess, however, mitigates the gender gap in knowledge but does not eliminate it completely (see also [Lizotte and Sidman, 2009](#); [Ferrín, Fraile, and García-Albacete, 2017](#)). Other aspects of the survey context have been shown to affect gender differences in political knowledge as well. [McGlone, Aronson, and Kobrynowicz \(2006\)](#) present evidence that the gender gap is exacerbated in an environment that induces stereotype threat, such as if women are aware of the fact that the study focuses on gender differences or if they are interviewed by a male interviewer. However, gender differences are not only induced by *how* researchers ask their questions, but also by the question *content*. Focusing on gender-relevant political knowledge items such as information about women's representation in the federal government has been shown to close the gap ([Graber, 2001](#); [Dolan, 2011](#); [Fraile, 2014b](#); [Jerit and Barabas, 2017](#)). Similarly, the gender gap is reduced or disappears when people are asked about more practical issues related to the government such as the availability of benefits and services ([Stolle and Gidengil, 2010](#)), or in political contexts characterized by more equitable representation of women ([Pereira, 2019](#); [McAllister, 2019](#); [Wolak and Juenke, 2019](#)). Importantly, women's lower factual knowledge score does not appear to impede on their political competence. In fact, [Dassonneville et al. \(2020\)](#) find that women are no less likely to vote for candidates who represent their preferences, and are therefore able to participate in politics just as effectively as men.

Overall, the gender gap appears to be influenced by how we ask for political information in surveys, as well as the kind of knowledge that is required for a correct response. Indeed, a comprehensive cross-national analysis of election studies in 47 countries between 1996 and 2011 suggests that question format and content account for large portions of the variance of gender disparities in political knowledge ([Fortin-Rittberger, 2016, 2020](#)). In short, conventional knowledge measures have problematic measurement properties that

may exacerbate observed gender differences.

Back to the Roots: The Structure of Belief Systems

Despite the discipline’s reliance on off-the-shelf item batteries, factual knowledge about political institutions has little overlap with the requirements for competent decision-making in politics—leading some scholars suggest that we should abandon recall questions as measures of “quality opinion” (Druckman, 2014). From a theoretical perspective, factual knowledge is all but a proxy for an underlying latent trait—political sophistication—which is usually conceptualized based on people’s belief systems instead of focusing on isolated pieces of factual information stored in declarative memory. 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).

How these ideas and attitudes—or considerations—are ultimately structured can differ along three different dimensions (Luskin, 1987). The first, and most obvious one, is the *size* of a belief system, which simply describes the number of distinct considerations that are available for retrieval. Politics, however, is comprised of a diverse set of independent domains—with some people having a deep grasp of a narrow field and others having a broad and potentially more shallow understanding of various issues. Thus, the second dimension describes the *range* of a belief system across domains (e.g., different policy issues or other evaluative categories). The last dimension is a belief system’s *constraint*, which describes the extent to which considerations are organized in a meaningful way through differentiation and integration of competing cognitions (Luskin, 1987). In other words, this dimension captures whether available considerations are perceived as operating in isolation or are rather as part of a more complex interconnected system, for example by identifying inherent value conflicts (Tetlock, 1983, 1993). To summarize, I conceptualize political sophistication based on the *structure* of individual belief systems

along the following three dimensions:

1. **Size:** *The number of considerations associated with a given category or issue.*
2. **Range:** *The dispersion of considerations across different categories or issues.*
3. **Constraint:** *The extent to which considerations are interconnected in a meaningful way.*

Political sophistication, in turn, is 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). Similarly, Tetlock (1983, 1993) coined the term *integrative complexity* to describe the degree to which considerations related to an issue are interconnected. In short, sophisticated political reasoning should reflect this notion of complex belief systems.³

Measuring Discursive Sophistication

Given that recall questions are only an imperfect measure for political sophistication, it is worth considering alternative—and potentially more imminent—observable implications of the underlying latent trait of interest: complex and highly constrained political belief systems. In the following, I propose a framework that leverages the content of open-ended responses in conjunction with the survey structure to evaluate how people discuss their political beliefs and preferences in their own words. To illustrate the approach in the context of a concrete example, consider a questionnaire where respondents is prompted to answer the following open-ended prompt:

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.

³It should be no surprise that Converse and others examined open-ended responses in their early studies—albeit from a slightly different perspective than the approach outlined here. Importantly, instead of relying on manual coding of open-ended responses, I develop an automated framework that is easily reproducible and can directly be applied to large surveys.

In addition, the survey includes similar questions on other topics such as **abortion**, **immigration**, **health care**, and **trade policies**—each asking respondents for both positive or negative considerations related to specific policy proposals. How would a complex and constrained set of political beliefs manifest itself across such a battery of open-ended responses? I argue that each dimension outlined above has direct observable implications for individual response behavior.

First, the *size* of a belief system is defined as the number of available considerations associated with a given category or issue. In the context of open-ended survey questions, a large belief system should therefore allow people to discuss their views by raising a larger number of distinct topics in response to each query. While this could also be achieved through manual coding, I rely on the structural topic model framework to extract the number of topics mentioned by each respondent in a survey (Roberts et al., 2014).⁴ Let \mathcal{W}_i denote the set of words contained in a response of individual i . Each word $w \in \mathcal{W}_i$ is assigned to a topic $t^* \in \{1, \dots, T\}$, such that $P(t^*|w, X_i) > P(t|w, X_i) \forall t \neq t^*$.⁵ In other words, each unique term in a response is assigned to the topic that has the highest likelihood of having generated that term, given the model. The set of topics that are mentioned by respondent i across all words in \mathcal{W}_i can then be described as \mathcal{T}_i^* and the number of considerations can be written as:

$$\text{size}_i = \frac{|\mathcal{T}_i^*|}{\max |\mathcal{T}_i^*|}. \quad (1)$$

I re-scale the measure to range from zero to one by dividing raw count of topics by the

⁴Please refer to the appendix for additional information. Specifically, see Appendix B for descriptive information on open-ended responses in each data set and structural topic model results. Appendix I contains further details on pre-processing steps and modeling choices for the structural topic models as well as robustness checks, which include preText analyses proposed by Denny and Spirling (2018).

⁵Note that $P(t|w, X_i) = \frac{P(w|t)P(t|X_i)}{P(w|X_i)}$. In the context of structural topic models, X_i denotes the covariates used to predict individual topic prevalence (see Roberts et al., 2014, for details). I used measures for age, gender, education, party identification, as well as an interaction between education and party identification as covariates for topic prevalence. This variable selection—with the exception of including gender—is equivalent to the procedure described in Roberts et al. (2014).

maximum number of topics observed across individuals.

Second, the *range* of a belief system is defined as the dispersion of considerations across categories or issues. Given a set of survey prompts covering various political issues, high levels of sophistication should correspond with people's ability to respond to each query with comparable levels of elaboration. I therefore quantify the consistency in response behavior across items by computing the Shannon entropy in open-ended response lengths:

$$\text{range}_i = \frac{-\sum_{j=1}^J p_{ij} \ln p_{ij}}{\ln J} \quad (2)$$

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

The last component addresses the level of *constraint* between considerations. The extent to which considerations are interconnected in a meaningful way should be associated with people's ability to differentiate and/or integrate them in their reasoning (Tetlock, 1993). Following Tausczik and Pennebaker (2010), I rely on specific function words as linguistic markers for these processes. More specifically, differentiating competing considerations in speech is usually accomplished using exclusive words (e.g., but, without), while integrating multiple thoughts is accomplished by the use of conjunctions (e.g., and, also). Thus, I measure relative constraint by identifying the number of conjunctions (CONJ_i) and exclusive words (EXCL_i) in each open-ended response using the Linguistic Inquiry and Word Count (LIWC) dictionary (Pennebaker et al., 2015):

$$\text{constraint}_i = \frac{\text{CONJ}_i + \text{EXCL}_i}{\max [\text{CONJ}_i + \text{EXCL}_i]} \quad (3)$$

As before, I re-scale the measure to range from zero to one by dividing all values by the empirical maximum observed across all individuals in the data.

Together, the three measures can be combined in a composite metric of sophistication in political attitude expression by averaging them for each respondent. Like each individual component, the resulting *discursive sophistication* score ranges from 0 to 1:

$$\text{discursive sophistication}_i = \frac{1}{3}(\text{size}_i + \text{range}_i + \text{constraint}_i). \quad (4)$$

Overall, a highly sophisticated individual should therefore give a more *elaborate* response across the full *range* of questions by *integrating* and/or *differentiating* multiple considerations. Note that this simple framework makes no assumptions about the direction of people's attitudes or their specific ideology. Crucially, since it is solely based on *how* individuals discuss their preferences, it can be directly applied in various settings to target specific political issues or tasks such as choosing between candidates running for election.

Of course, this is not the first time a framework is developed to assess the complexity of written (or spoken) word. In fact, this task has been the subject of longstanding research in linguistics and educational sciences, resulting in a multitude of alternative metrics. Recently, these measures have been employed by political scientists who study different forms of elite communication. [Spirling \(2016\)](#), for example, uses a standard readability score based on the ratio of words per sentence and syllables per word to study the linguistic complexity of speeches in the British House of Commons over time. More recently, [Benoit, Munger, and Spirling \(2019\)](#) expanded on previous metrics to develop a measure of comprehensibility that is more applicable in the realm of politics.

These approaches—and especially the development of metrics specifically suited for political text—are particularly useful when studying elite communication. Yet, in contrast to the framework outlined above, they focus on the *comprehensibility* as a measure of complexity; elite sophistication is evaluated based on a recipient's ease to understand the message, which is largely driven by linguistic and syntactic difficulty rather than actual political content. While this is certainly a reasonable approach when studying the effects

of elite communication, the inference of interest outlined in this paper is markedly different. My focus is to examine verbatim attitude expression to assess the underlying degree of elaborate political reasoning. Pure linguistic style is therefore not of central concern so long as it is unrelated to the actual political content.⁶ After all, being hard to comprehend does not necessarily imply that someone put a lot of thought into a statement.

Results

Throughout the remainder of this paper, I apply the proposed measure of discursive sophistication across multiple surveys that employed different sets of open-ended questions. Below is a short summary of the data sets and items:⁷

- **Cooperative Election Study (CES 2018):** 10 open-ended questions targeting policy preferences on gun legislation, abortion, immigration, health care, and trade.
- **American National Election Study (ANES 2020, 2016, 2012):** 8 open-ended likes-dislikes questions targeting preferences for parties and candidates.
- **YouGov Survey (2015):** 4 open-ended questions targeting policy preferences on gun legislation and health care.
- **Swiss Referendum Surveys (2012-2008):** 2 open-ended questions asking respondents to justify their vote choice in various policy referenda. These surveys were conducted in three languages (French, German, and Italian).

I proceed by providing descriptive evidence regarding the face validity of discursive sophistication. Next, I assess its construct validity by comparing it to factual knowledge as a predictor of various relevant outcomes such as political participation and engagement. The last validation step consists of comparing discursive sophistication to manually coded levels of justification in open-ended responses. Each of these steps leverage different subsets of the studies listed above, depending on the availability of necessary

⁶In fact, pure linguistic complexity is arguably driven more by other factors such as a person's general verbosity or linguistic prowess and therefore less valid as a measure of political sophistication.

⁷A detailed description of each data set and the specific question wording is included in Appendix A.

items. After validating the measure, I assess gender gaps in discursive sophistication and factual knowledge using the complete set of surveys.

A First Look at Discursive Sophistication

While each dimension of discursive sophistication outlined above provides a unique source of variance to the underlying concept (Luskin, 1987), all three are positively correlated.⁸ Furthermore, exploratory factor analyses confirm that they load on a single factor with all loadings exceeding 0.5 across the CES and ANES data—thus confirming that we can rely on an additive score to measure discursive sophistication.⁹ Detailed results are presented in Table 1.

Variable	2018 CES	2020 ANES	2016 ANES	2012 ANES
Size	0.840	0.997	0.997	0.997
Range	0.526	0.536	0.548	0.576
Constraint	0.513	0.684	0.623	0.709

Table 1: Factor Loadings of Discursive Sophistication Components

How does this discursive sophistication score compare to alternative metrics of political knowledge? As discussed in the beginning, the standard approach to measuring political knowledge in surveys is to ask a set of factual questions about political institutions. The CES and ANES include 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 responses on these items to form an additive index of *factual knowledge* about politics. As an additional benchmark, I consider *interviewer assessments* of each respondent’s political sophistication (see Bartels 2005; but cf. Ryan 2011).¹⁰

Figure 2 compares discursive sophistication to the conventional knowledge metrics

⁸See Appendix B for correlation matrices between individual components.

⁹I rely on the CES and ANES here since these surveys employ a larger set of open-ended questions.

¹⁰Note that interviewer assessments were only recorded in the face-to-face sample of the 2012 and 2016 ANES.

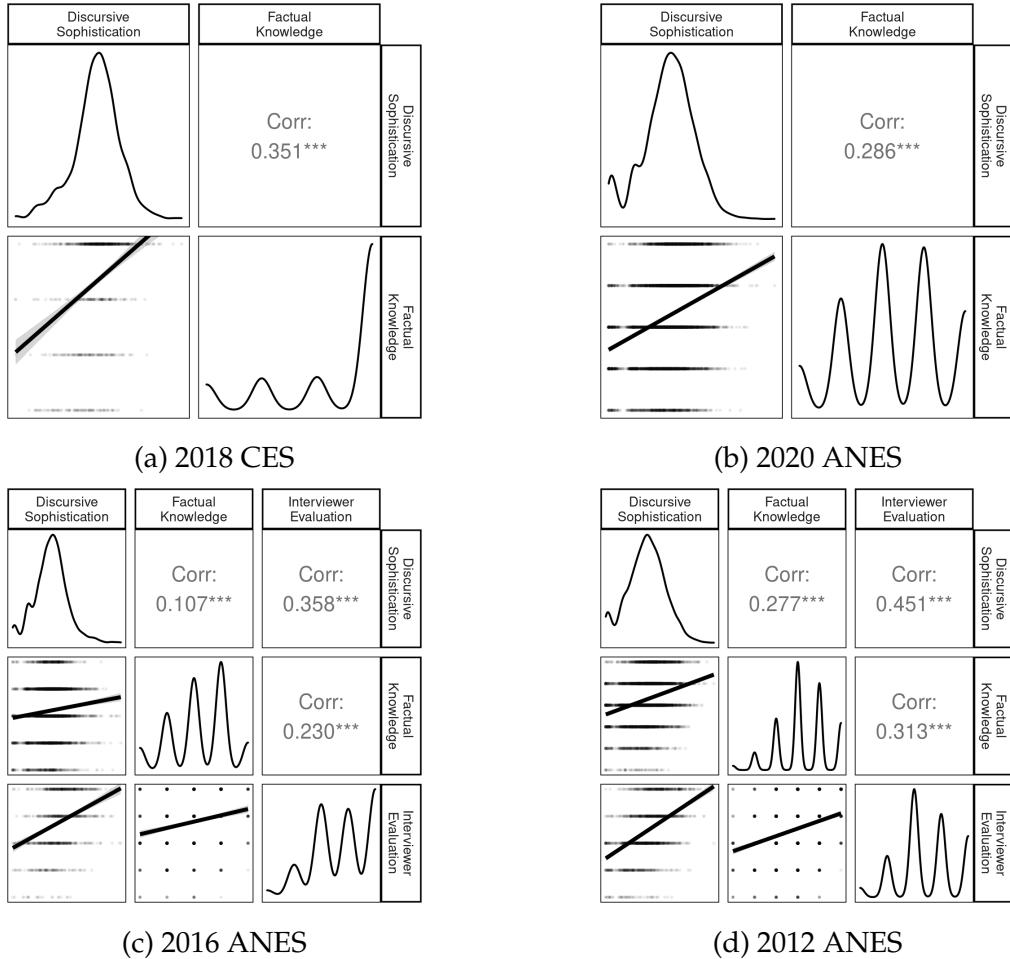


Figure 2: Correlation matrix of discursive sophistication and conventional political knowledge metrics. The plots on the diagonal display univariate densities for each variable. The panels in the lower triangular display the scatter plot of two measures as well as a linear fit. The upper triangular displays the correlation coefficient. All correlations reported are statistically significant with $p < .05$.

for the CES and ANES. Each figure presents scatterplots between individual measures (lower triangular), univariate densities (diagonal), and correlation coefficients (upper triangular). The measure of discursive sophistication is positively correlated with both conventional metrics while capturing some additional variation. Interestingly, there is a stronger correlation between discursive sophistication and interviewer evaluations than between factual knowledge and interviewer evaluations ($r = .36$ vs. $r = .23$ in 2016, and $r = .45$ vs. $r = .31$ in 2012), which indicates that the open-ended measure captures characteristics that influence subjective assessments of sophistication. Thus, a respondent's

verbatim answers seem to be more influential for subsequent knowledge assessments by the interviewer than a respondent's performance on the factual knowledge questions.

While discursive sophistication and the alternative measures are clearly correlated, the relationship between each metric is far from perfect. To provide some intuition as to whether the variation in discursive sophistication is theoretically meaningful, I present an example of open-ended responses of two individuals in the 2018 CES who *scored equally on factual knowledge* (3 out of 3 correct responses), but varied in discursive sophistication.

	A: Low Sophistication Response	B: High Sophistication Response
Guns (+)	Increases the fact that bad actors will be caught before getting guns to cause problems	Making sure healthy law abiding citizens only can get guns
Guns (-)	Can't think of any	The second amendment....
Abortion (+)	Don't know	Killing babies
Abortion (-)	Don't know	Women want the right to decide what to do medically with their own bodies
Immigration (+)	It was not their fault they were brought here. If they are not committing crimes, they should be given a chance at citizenship	They have been living here for years and should not have to be uprooted
Immigration (-)	don't know	They came here illegally and are criminals
Health care (+)	It is too expensive - it requires people who don't use insurance or haven't in the past to pay for everybody who does, especially the people who are getting it for free.	Cost to much, hurts the middle class, raised costs of health insurance
Health care (-)	Can't think of a thing.	Many people who get insurance through the ACA would be left uninsured
Trade policy (-)	don't know	More money for the government without taxing the people
Trade policy (+)	don't know	They are affecting trade relations, making products brought into the us more expensive
Disc. Soph.	0.289	0.517

Table 2: Example of open-ended responses for low and high scores on discursive sophistication with equal factual knowledge scores (3 out of 3 correct responses). Column A displays the verbatim responses of an individual who scored low on discursive sophistication and column B displays the verbatim responses of an individual who scored high on the open-ended measure. Each row represents one of the likes/dislikes items included in the analysis. Note that the entries are original responses without editing.

The results are presented in Table 2. Each row represents one of the open-ended re-

sponses targeting specific policy issues. Column A displays the responses of an individual who scored low on discursive sophistication and column B displays the responses of a high scoring individual. Even though both individuals have the same factual knowledge score, there are systematic differences in their response behavior that suggest disparity in their political sophistication. Overall, respondent A provided a less elaborate response and only focused on a narrow range of issues. Irrespective of whether one agrees with the specific statements, A's response pattern is suggestive of a less sophisticated political belief system and a lower level of motivation to engage in in-depth reasoning about each issue. Overall, this initial result suggests that the variation in discursive sophistication captures meaningful differences in response behavior that overlaps with traditional knowledge metrics while displaying some unique variation. The following sections will show that this variation is also politically consequential.

Validating the Measure

A crucial step in validating any measure of political sophistication is to examine the extent to which it is correlated with political engagement and citizen competence (Lupia, 2006, 2015). Accordingly, I consider how discursive sophistication is associated with (1) engagement and participation in politics, (2) the ability to incorporate new information, and (3) well-justified policy preferences.

Engagement and Participation in Politics

Any measure of political sophistication should be strongly associated with individual engagement and participation in politics. In fact, factual knowledge items have been validated in the past based on their strong relationship with outcomes such as turnout and other forms of participation (Lupia, 2015, 230–233). Figure 3 compares the effect of discursive sophistication and factual knowledge on four dependent variables related to political engagement: turnout, political interest, internal efficacy, and external efficacy. The model

predicting turnout is estimated via logistic regression while the estimates for the three remaining dependent variables are based on OLS. Each model controls for gender, education, income, age, race, and church attendance.¹¹

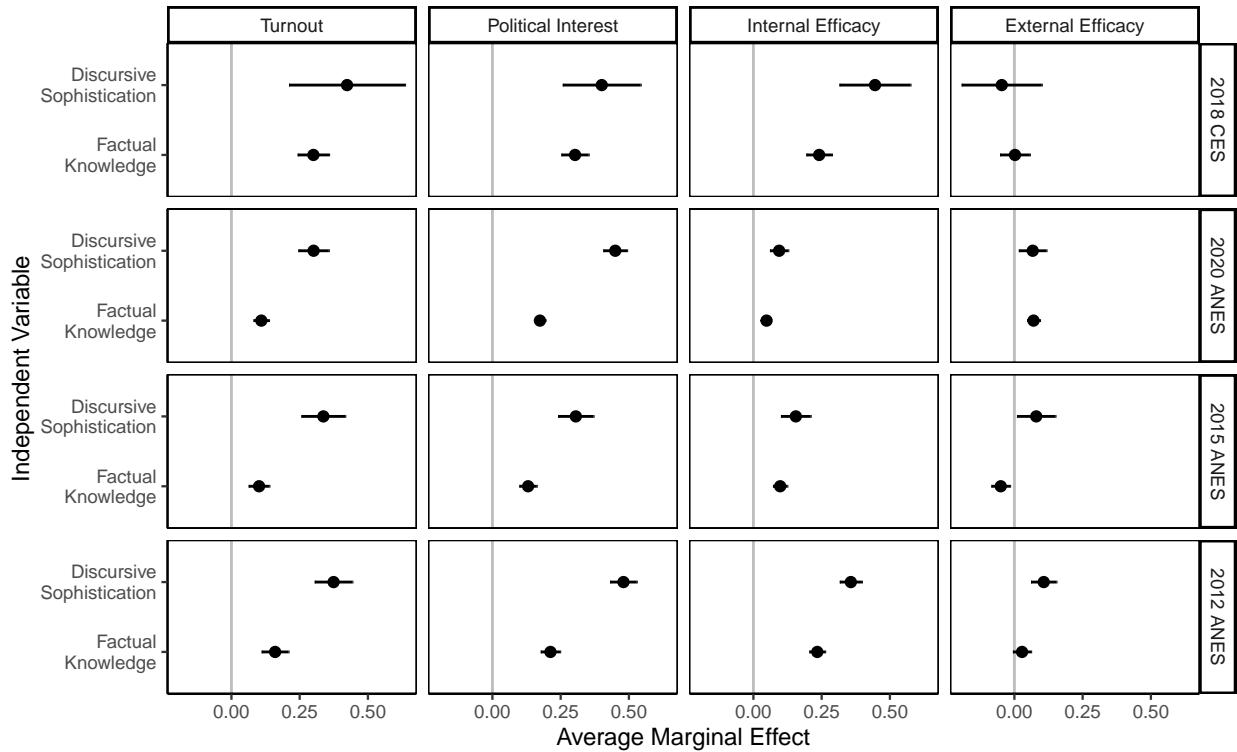


Figure 3: Effects of sophistication on turnout, political interest, internal efficacy, and external efficacy in the CES and ANES. For each dependent variable, the figure displays the average marginal effects (AME) for each sophistication measure (including 95% confidence intervals). Model estimates are based on logistic regression (turnout) or OLS (political interest, internal efficacy, external efficacy). Each analysis includes controls for gender, education, income, age, race, and church attendance. Full model results are displayed in the appendix, Tables D.1 through D.4.

Each panel compares the average marginal effects of both sophistication measures on the respective dependent variable while holding all other variables constant at their means. Of course, these effects are purely correlational and should not be interpreted causally. Nevertheless, across all four surveys, discursive sophistication is a stronger predictor of turnout, political interest, and internal efficacy. The results for external efficacy

¹¹Appendix A provides additional information on these as well as remaining variables included in subsequent analyses.

are more ambiguous. Factual knowledge has strikingly inconsistent effects—sometimes predicting higher, lower, or no change in external efficacy. Discursive sophistication, in contrast, is more consistently associated with higher external efficacy (the only exception is the 2018 CES, which uses a shorter battery to measure external efficacy).

Considering these initial results, a potential concern may be that discursive sophistication is confounded by individual characteristics that influence verbatim response patterns as well as engagement. Appendix III provides additional analyses controlling for factors that might drive verbosity such as personality (extraversion), survey mode (face-to-face vs. online), verbal skills, as well as individual response length itself. The substantive conclusions remain unchanged.

Incorporation of New Information

In order to replicate and extend this first validation, I rely on a separate nationally representative survey employing an alternative set of open-ended responses. The data was collected by YouGov in December 2015 and contains responses of 1000 U.S. citizens. As part of this study, respondents were asked four open-ended questions to describe their attitudes towards two salient issues: gun legislation and the Affordable Care Act.

Political sophistication should imply the ability to incorporate relevant new information about parties, office-holders, and policies. After all, Zaller (1990, 1992) and others argue that tests of factual information about politics are the best available proxy for awareness. In this analysis, I explore whether discursive sophistication or factual knowledge serves as a better predictor of people's ability to incorporate new information from media sources. As part of the survey, respondents were asked to read a newspaper article about a fictional infectious disease and were subsequently asked to answer questions about 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 (*information retrieval*) 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.

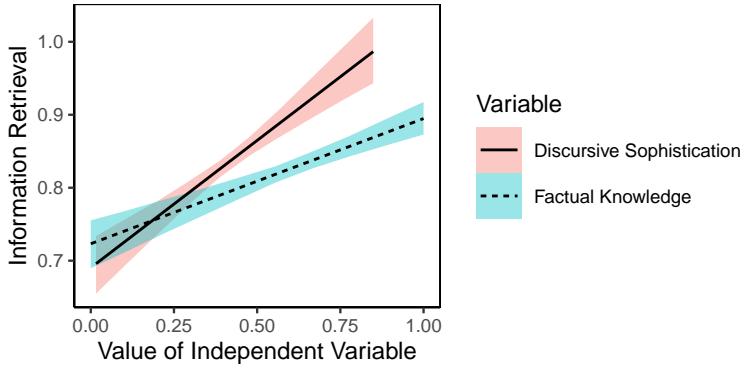


Figure 4: Expected information retrieval in the 2015 YouGov Study as a function of political sophistication (including 95% confidence intervals). Estimates are based on a linear regression model controlling for education, income, age, church attendance, gender, and race. Full model results are displayed in the appendix, Table D.5.

Figure 4 displays the relationship between political sophistication and disease information retrieval in the 2015 YouGov study. Estimates are based on linear regression models controlling for education, income, age, church attendance, gender, and race. As a benchmark for discursive sophistication, I again consider the effect of factual knowledge based on a battery of eight items similar to the knowledge questions in the ANES. Both discursive sophistication as well as factual knowledge increase the amount of information individuals are able to recall from a news article discussing a fictional disease. Similar to the previous results, the effects are stronger for discursive sophistication than for factual knowledge scores. The degree to which citizens discuss their own political beliefs in a more elaborate manner is not only a stronger predictor of political engagement but also serves as a better proxy for the ability to incorporate new information about a non-partisan issue.

Well-Justified Policy Preferences

As the last validation step, I examine an additional set of surveys that provide a unique opportunity to compare my proposed measure of discursive sophistication with manu-

ally coded open-ended responses across three languages. Colombo (2016) compiled a data set of cross-sectional surveys administered in Switzerland after national popular votes on multiple policy propositions. For each referendum, respondents were asked to explain why they voted in favor or against a given proposition in two separate open-ended items. Based on these verbatim responses, I computed the discursive sophistication using the same procedure outlined above. Since the survey was conducted in three different languages (German, French, and Italian), I created separate metrics for each group of respondents.

Beyond the ability to incorporate new information, political sophistication should enable people to justify their own preferences. Colombo's (2016) manual coding of the respondents' *level of justification* assessed the content, elaboration, and complexity of open-ended responses. Thus, this study provides an opportunity to directly assess the extent to which high levels of discursive sophistication correspond to well-justified policy preferences in open-ended responses. Any overlap between Colombo's (2016) manual coding with my automated measure corroborates the face validity of discursive sophistication.

The results are presented in Figure 5, which displays the distribution of discursive sophistication for each level of justification coded by Colombo (2016) as well as the correlation coefficients for both respective variables. Across all three language groups, discursive sophistication is systematically higher among respondents with the highest level of justification and both measures are positively correlated ($r = 0.23, 0.32$, and 0.36 , respectively). The proposed measure of discursive sophistication therefore shows a high degree of correspondence with individual levels of justification assessed by independent manual coders.

To summarize, the results presented thus far indicate that discursive sophistication shares common characteristics with factual political knowledge measures. At the same time, the proposed measure outperforms conventional metrics as a predictor of political participation and engagement, provides a better proxy for the ability to incorporate

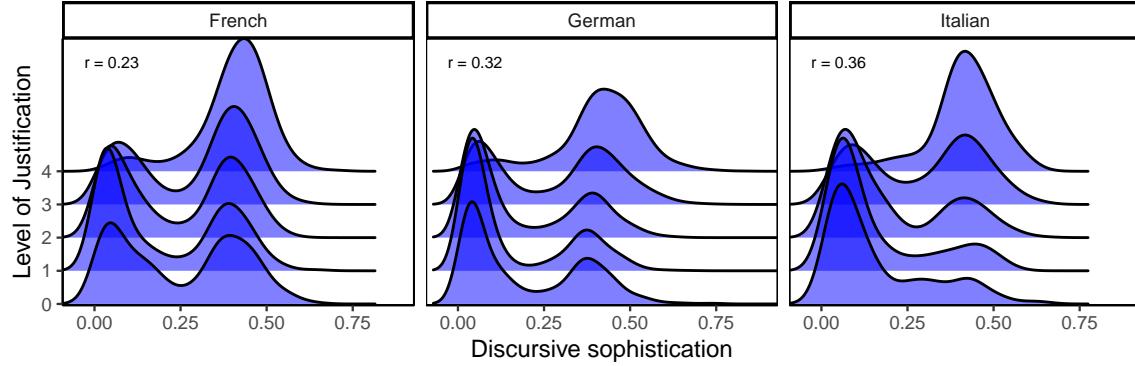


Figure 5: Discursive sophistication and manually coded level of justification (Colombo, 2016) in Swiss post-referendum surveys. The plot compares kernel densities of discursive sophistication for each manually coded level of justification.

new information from news sources, and shares significant overlap with manually coded levels of justification in open-ended responses. In the following, I illustrate how discursive sophistication can help refine previous findings regarding the gender gap in political knowledge.

Reassessing the Gender Gap

How do women and men compare on the different metrics of political sophistication in the surveys analyzed in the present study? Figure 6 displays the average levels of discursive sophistication and conventional metrics comparing both genders. While we observe a sizable and statistically significant gender gap for factual knowledge across the CES, ANES, and YouGov surveys, this difference disappears for discursive sophistication. Even though women do not perform as well as men on political quizzes, they do not differ substantially in complexity and sophistication when describing their political preferences.

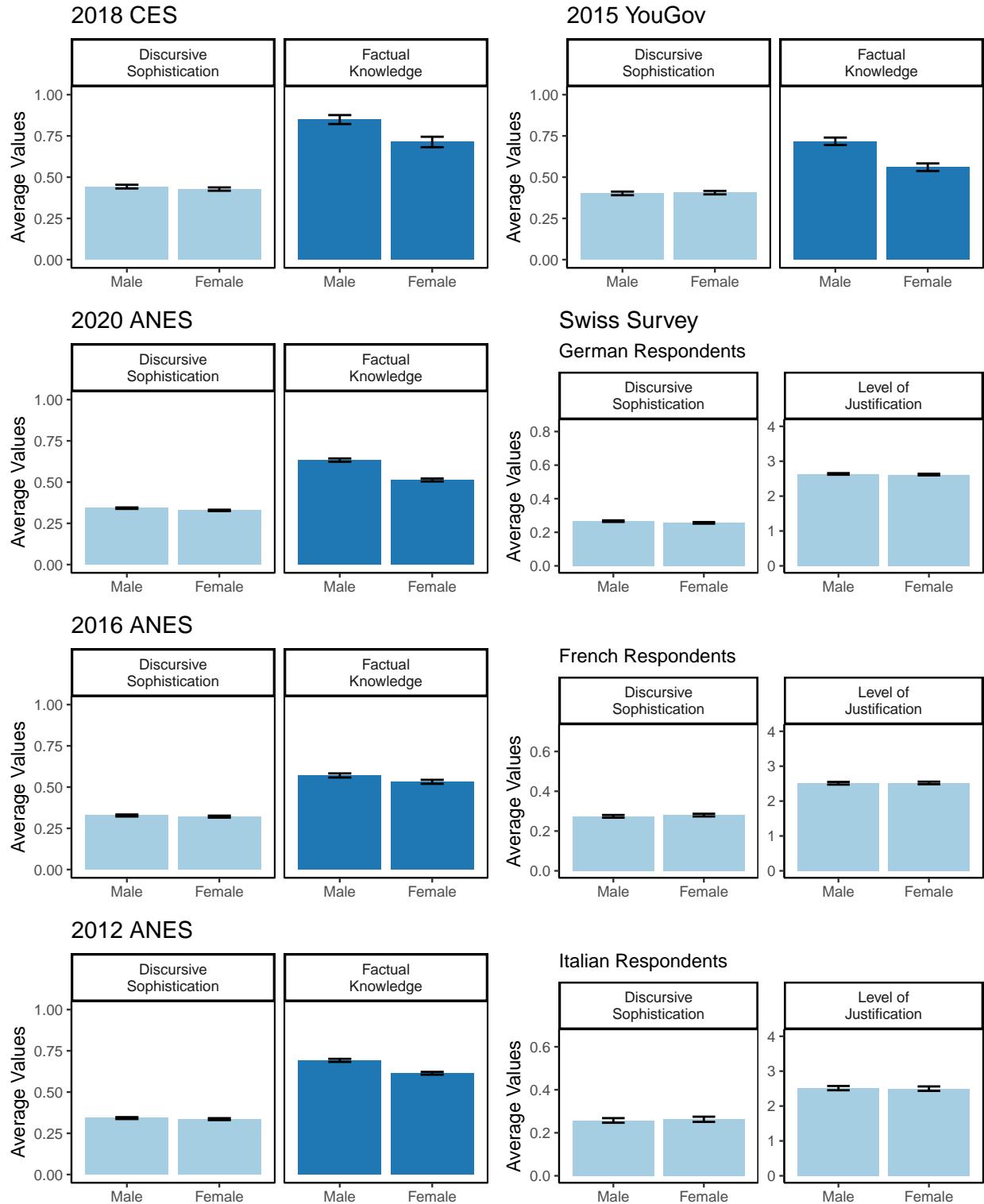


Figure 6: The gender gap in political sophistication. The figures display mean levels of discursive sophistication (light blue) and factual knowledge (dark blue) comparing women and men (including 95% confidence intervals). Gender differences in factual knowledge in the CES, ANES, and YouGov survey are statistically significant with $p < .05$.

Of course, we need to make sure that this absence of a gender gap in discursive sophistication is not idiosyncratic to the particular measurement approach proposed here. One way to investigate this question is to examine gender differences in discursive sophistication using data from Colombo (2016) and comparing them to her manually coded measure. That way, we can not only determine whether the lack of a gender gap in discursive sophistication replicates in the Swiss survey, but also check whether there is an equivalent lack of gender differences in Colombo's alternative measure of citizen competence in direct democracies. If discursive sophistication captures a person's motivation to undertake in-depth reasoning and form quality opinions (and assuming these characteristics do not differ by gender), there should be no difference between women and men on either metric (discursive sophistication and Colombo's measure). As shown in the bottom row of Figure 6 there are indeed no significant gender differences on *both* metrics across all three languages in the Swiss referendum surveys. The absence of a gender gap is consistent whether open-ended responses are coded manually or using the proposed measure of discursive sophistication.

Next, we have to consider whether the apparent gender gap in factual knowledge is a manifestation of real differences between women and men. Prior research attributes at least part of the gap to actual discrepancies in individual resources and engagement. Accordingly, we need to control for these determinants of political knowledge to provide a more comprehensive examination of the veracity of observed gender differences. Figure 7 displays estimated gender differences after controlling for various potential common determinants such as education, income, age, race and church attendance.

After controlling for common determinants, discursive sophistication again reveals no significant differences between women and men across the CES, ANES and YouGov surveys.¹² The gender gap in factual political knowledge, however, persists and is sub-

¹²I did not include the Swiss study in this comparison because the survey did not include comparable factual knowledge items. Suffice to say that there are no significant gender differences in discursive sophistication in the Swiss study when including sociodemographic controls.

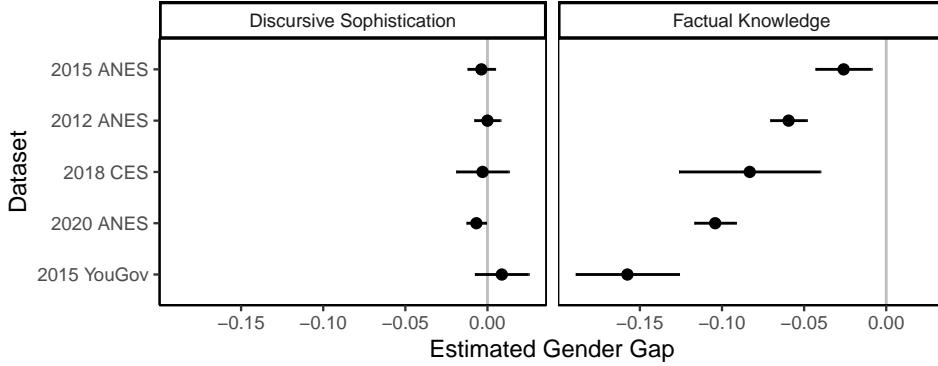


Figure 7: The gender gap in political sophistication controlling for common determinants. Estimates are OLS regression coefficients with 95% confidence intervals. Dependent variables are discursive sophistication and factual political knowledge. Estimates are based on a linear regression model controlling for education, income, age, church attendance, gender, and race. Full model results are displayed in the appendix, Tables D.6 and D.7.

stantively as well as statistically significant. Thus, a considerable portion of the observed differences in factual knowledge between women and men cannot be attributed to underlying disparities in resource-related factors or engagement. Comparing the confidence intervals across both measures further reveals that the insignificant gender differences in discursive sophistication are estimated with higher precision than the significant differences in factual knowledge. Such a result precludes the possibility that null findings for discursive sophistication are purely driven by measurement error on the dependent variable. It is also worth pointing out in this context that the effects of control variables are quite similar across both measures and different surveys.¹³ For instance, knowledge and discursive sophistication are significantly higher among respondents who are more educated and have higher income. The finding that determinants of political sophistication are consistent across models lends additional validity to the open-ended measure.

Explaining the (Lack of a) Gender Gap

To summarize, conventional knowledge measures and discursive sophistication produce diverging conclusions regarding the existence of a gender gap. This naturally raises the

¹³See Appendix D for full regression results.

question which metric we should ultimately trust? Prior research attributed gender differences in factual knowledge—at least partly—to the format (e.g., availability of “Don’t Know” options) and content (e.g., focusing on issues that are less relevant to women) of item batteries. This section explores whether these arguments provide a sufficient explanation for the conflicting results for discursive sophistication—namely the complete lack of systematic differences between women and men. In other words, which one is more likely to be an artifact of the respective measurement approach: the *existence* of a gender gap in factual knowledge or the *absence* of a gap in discursive sophistication?

The first set of arguments about why conventional metrics may overstate potential gender differences is based on the finding that women are less likely to guess than men (Mondak and Anderson, 2004). Arguably, respondents’ differential willingness to admit not knowing the answer to a question is certainly less of an issue when they are simply asked to voice their opinions rather than being quizzed on political facts. Following best practices, however, the surveys presented here omitted “Don’t Know” options in their recall questions. Differential propensity to guess can therefore not be viewed as a valid explanation for the gender gap in factual knowledge observed here. At the same time, the lack of significant differences between women and men in discursive sophistication may itself be the product of selection biases in women’s willingness to answer open-ended questions in the first place. Following this argument, it could be the case that only women who are highly sophisticated provide a response, thereby misleadingly closing the gender gap in the discursive measure. There are two reasons why that is unlikely to be the case. First, as the analyses have shown, this proposed selection mechanism does not diminish gender differences in factual knowledge. Second, and more importantly, there are no significant differences between men’s and women’s willingness to answer open-ended questions. In fact, adjusting for potential selection effects when examining determinants of sophistication does not change the substantive conclusions.

The second major explanation for the gender gap in political knowledge focuses on

the question content. By choosing a specific set of recall questions as a general metric for political knowledge, researchers are making strong assumptions about the information deemed necessary for competent decision-making. As it turns out, these item batteries usually focus on male-dominated topics in politics (Dolan, 2011). Open-ended questions, on the other hand, make it possible to directly study the information that is in fact available to citizens and—importantly—to examine how they apply their knowledge when discussing their political preferences.

Accordingly, if it is the case that the gender gap in discursive sophistication is nonexistent simply because open-ended questions allow women to raise political considerations particularly salient to them, then we should be able to observe systematic variation in types of issues discussed by women and men, respectively. Luckily, we can directly examine such gender differences in topic prevalence within the structural topic model framework used to measure discursive sophistication. More specifically, gender is included in the model as one of the covariates that influences how often each topic is discussed by a respondent (see also Roberts et al., 2014, for details).

In this last analysis, I therefore explore how women and men differ in topical prevalence across open-ended responses in the 2012, 2016, and 2020 ANES. Note that these open-ended items did not focus on specific issue areas as in the CES, but rather asked respondents to evaluate different political parties and candidates. Thus, they were able to focus on whatever issue they deemed most important. Figure 8 displays the subset of topics that shows the largest absolute gender difference in topic prevalence in both waves. Positive coefficients indicate that women are more likely than men to mention a given topic, and vice versa. The top five topics are more prevalent among men and the bottom five are more likely to be mentioned by women. The label for each coefficient consists of the five highest probability terms related to the topic to illustrate its content.

Taking the 2012 ANES as an example, the topic consisting of terms such as *care*, *health*, and *reform* is significantly more likely to be mentioned by women. On the other hand,

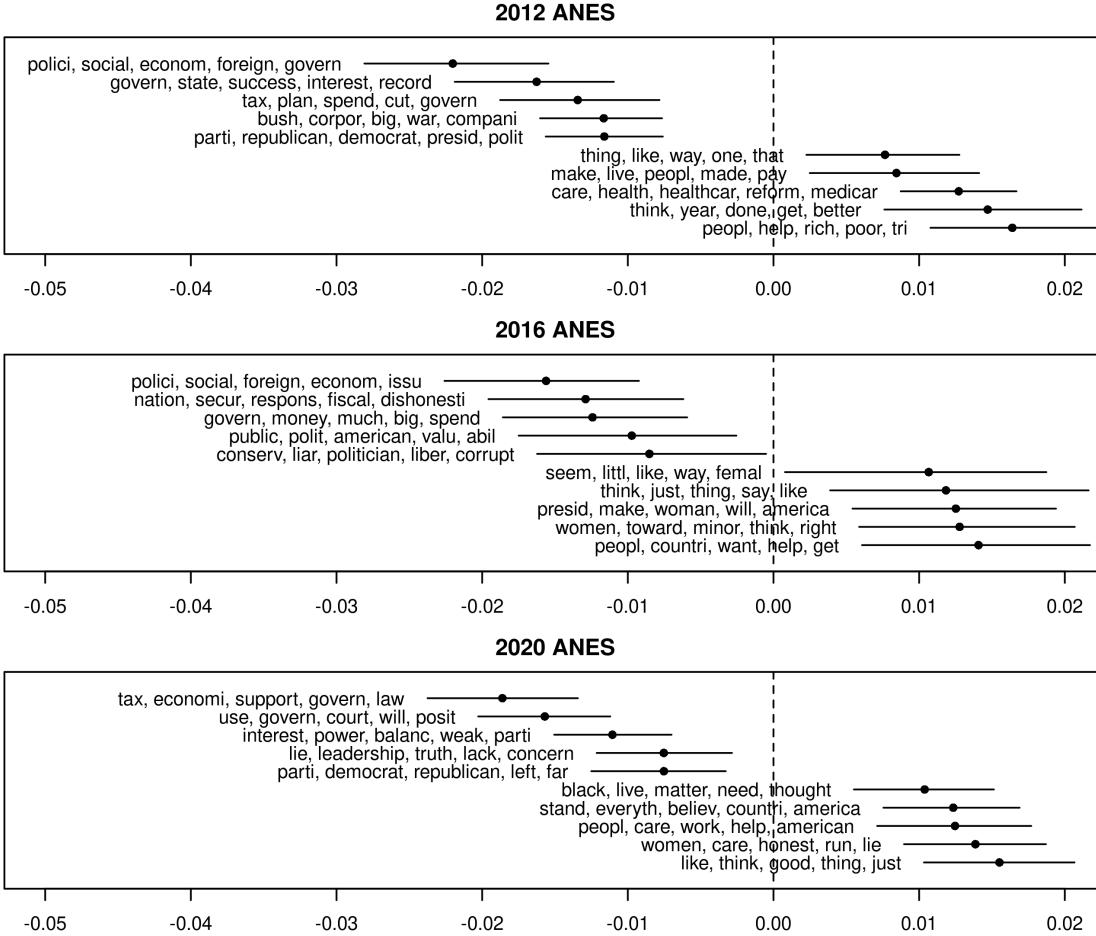


Figure 8: Gender differences in topic proportions in open-ended responses based on the structural topic model used to compute discursive sophistication (including 95% confidence intervals). Coefficients indicate the difference in predicted topic prevalence among women and men; positive values indicate higher prevalence among women. Labels are based on the five highest probability terms related to the topic.

men are more likely to mention the topic revolving around terms like *tax*, *spend*, and *cut*. Overall, across both waves of the ANES, women were less likely than men to discuss foreign affairs, economic issues, or the Supreme Court. Instead, they focused on issues related to women's rights, equality, and health care. The considerations raised by women when discussing their political preferences are therefore clearly different from men's and—crucially—the issues discussed by men happen to be more aligned with the type of questions usually covered in standard political knowledge batteries (i.e., pertaining to the economy, institutions, elites, etc.). For example, men are more likely to mention

considerations related to the federal budget in their open-ended responses. At the same time, two of the five knowledge questions included in the 2012 ANES pertain to government spending: one asking respondents to compare the federal deficit to levels in 1990, the other requiring a comparison of federal spending on different programs such as foreign aid, medicare, and national defense.

Overall, the results indicate that gender differences in conventional knowledge metrics are at least partly driven by the fact that the issues women care about are not represented in standard item batteries. When using the alternative measure—discursive sophistication—any evidence for systematic differences between women and men disappears since open-ended questions about political preferences allow respondents to focus on different issues.

Discussion

From a normative perspective, there is no reason to assume that a particular set of issues should be more important for citizens' preference formation or competence in elections. Whether one cares more about the federal budget or reproductive rights, the most important question is whether citizens think deeply about the issues they care about and incorporate them accordingly when making their vote choices. As Druckman (2014) argues, citizen competence (for example in elections) should not be evaluated based on their ability to recall unrelated facts about political institutions, but rather focus people's motivation to form quality opinions—which implies that they focus on the issues most important to them. As it turns out, while the types of issues raised women and men differ systematically, there is no reason to assume that women are therefore less sophisticated or competent in the realm of politics.

This issue has been recognized in the literature before (e.g., Graber, 2001; Dolan, 2011; Ferrín et al., 2020), but it cannot be properly addressed while relying exclusively on off-

the-shelf recall questions to measure political knowledge. What is more, there is thus far no principled approach to develop new sets of items that focus less on male-dominated issues. Beyond proposing an alternative measurement approach, the framework presented in this paper can help provide such a first step towards devising balanced recall items. More specifically, examining the types of issues women and men emphasize when discussing their political preferences can serve as a guide to select new sets of knowledge questions. Thus, future research should explore whether factual knowledge questions selected based on open-ended responses are indeed more balanced with regard to gender differences. To the extent that this proves to be a useful heuristic for item selection, researchers planning a survey could rely on pilot studies fielding open-ended questions in order to devise balanced factual knowledge items in the main survey.

That being said, relying on open-ended responses to assess political sophistication has its limitations. First and foremost, elaboration in verbatim attitude expression may be more prone to biases due to differential levels of motivation to answer survey questions. It should be noted, however, that conventional knowledge metrics are not free from survey effort effects either—as indicated for example by the fact that scores can be improved by providing monetary incentives for correct responses (Prior and Lupia, 2008)—and future studies should investigate the extent to which discursive sophistication is subject to similar deviations. A related potential confounding factor that is unique to open-ended responses is the respondents’ general linguistic skills or verbal verbosity, which may again influence elaboration in open-ended responses but is orthogonal to political sophistication.

One reason why these potential drawbacks may be less worrisome is that the proportion of respondents who refuse to answer any open-ended question in the first place is very low, which indicates that people are sufficiently motivated to engage with the survey. Furthermore, controlling for pure response length did not change the substantive conclusions regarding the effects of discursive sophistication on, for example, political

participation or efficacy. The results were also robust to the inclusion of measures of linguistic skills or personality characteristics such as extraversion. In a similar vein, the gender gap finding did not appear to be driven by selection effects, which again suggests that survey effort—albeit an important confounding factor to consider—is unlikely to jeopardize the substantive conclusions presented in this paper.

Lastly, even if one supports the general notion that open-ended responses can provide useful insights, a skeptic may still argue that manual coding is preferable to the automated framework presented here. However, manual coding of open-ended responses is not always feasible in the context of large-scale surveys, since it can be labor-intensive and requires extensive contextual knowledge such as high levels of language proficiency. The Swiss surveys in Colombo's (2016) study, for example, were conducted in three different languages (German, French, and Italian) and ranged across numerous policy referenda. More importantly, knowledge assessments can be biased by the level of political agreement between individuals (e.g., Ryan, 2011). The measurement approach presented here, on the other hand, is easily replicable and reproducible, is not affected by subjective judgments, and can be directly applied to large-scale surveys in multiple contexts across different languages.

Conclusion

Political scientists should worry less about pure levels of *factual knowledge* and instead focus on how people justify their political preferences. Factual knowledge about political institutions might be a useful proxy in certain scenarios, but it cannot address directly whether individuals hold well-considered opinions about political actors or issues. In comparison, the measure of discursive sophistication proposed here is agnostic about the specific contents of people's beliefs, but directly targets the complexity of expressed attitudes. It can therefore be easily applied to assess sophistication in any decision-making

context (such as policy referenda or local elections) by fielding targeted open-ended questions related to the relevant underlying beliefs and preferences. Furthermore, a free software package for the statistical programming environment R will allow applied researchers to implement the framework presented here.¹⁴

The findings presented in this paper show that conventional knowledge indices and the open-ended measure share a substantial amount of variance. However, they are far from being identical and capture different aspects of sophistication. In fact, discursive sophistication is a stronger predictor of political engagement and efficacy than traditional metrics. It is also strongly related to people's ability to incorporate new information from news sources and shows a high degree of overlap with manually coded levels of justification. Most importantly, using the discursive 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. Furthermore, the lack of gender differences in discursive sophistication can be attributed to the fact that open-ended questions allow women to focus on different considerations than men.

In the past, scholars have argued that testing for factual information, despite its shortcomings, still provides the best available measure of political awareness as it captures "what has actually gotten into people's minds, which, in turn, is critical for intellectual engagement with politics" Zaller (1992, 21). The results presented in this paper suggest that a direct examination of open-ended responses provides a viable supplemental approach that promises new insights into how people make up their mind about politics.

¹⁴Package under development, release on CRAN expected mid 2022.

References

- Abrajano, Marisa. 2014. "Reexamining the "Racial Gap" in Political Knowledge." *The Journal of Politics* 77 (1): 44–54.
- 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.
- Benoit, Kenneth, Kevin Munger, and Arthur Spirling. 2019. "Measuring and explaining political sophistication through textual complexity." *American Journal of Political Science* 63 (2): 491–508.
- Bos, Angela L, Jill S Greenlee, Mirya R Holman, Zoe M Oxley, and J Celeste Lay. 2021. "This One's for the Boys: How Gendered Political Socialization Limits Girls' Political Ambition and Interest." *American Political Science Review*: 1–18.
- Bullock, John G., and Kelly Rader. 2021. "Response Options and the Measurement of Political Knowledge." *British Journal of Political Science*: 1–10.
- 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* 80 (4): 858–887.
- Colombo, Céline. 2016. "Justifications and Citizen Competence in Direct Democracy: A Multilevel Analysis." *British Journal of Political Science*: 1–20.

- Converse, Philip E. 1964. "The Nature of Belief Systems in Mass Publics." In *Ideology and Discontent*, ed. David E. Apter. New York: Free Press.
- Cramer, Katherine J., and Benjamin Toff. 2017. "The Fact of Experience: Rethinking Political Knowledge and Civic Competence." *Perspectives on Politics* 15 (3): 754–770.
- Dassonneville, Ruth, Mary Nugent, Marc Hooghe, and Richard R Lau. 2020. "Do Women Vote Less Correctly? The Effect of Gender on Ideological Proximity Voting and Correct Voting." *Journal of Politics* 82 (3): 1156–1160.
- 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.
- 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.
- Dolan, Kathleen. 2011. "Do women and men know different things? Measuring gender differences in political knowledge." *The Journal of Politics* 73 (01): 97–107.
- Dow, Jay K. 2009. "Gender differences in political knowledge: Distinguishing characteristics-based and returns-based differences." *Political Behavior* 31 (1): 117–136.
- Druckman, James N. 2014. "Pathologies of studying public opinion, political communication, and democratic responsiveness." *Political Communication* 31 (3): 467–492.
- Ferrín, Mónica, Marta Fraile, and Gema García-Albacete. 2017. "The Gender Gap in Political Knowledge: Is It All about Guessing? An Experimental Approach." *International Journal of Public Opinion Research* 29 (1): 111–132.

Ferrín, Monica, Marta Fraile, Gema M García-Albacete, and Raul Gómez. 2020. "The gender gap in political interest revisited." *International Political Science Review* 41 (4): 473–489.

Fortin-Rittberger, Jessica. 2016. "Cross-National Gender Gaps in Political Knowledge How Much Is Due to Context?" *Political Research Quarterly* 69 (3): 391–402.

Fortin-Rittberger, Jessica. 2020. "Political knowledge: Assessing the stability of gender gaps cross-nationally." *International Journal of Public Opinion Research* 32 (1): 46–65.

Fraile, Marta. 2014a. "Do women know less about politics than men? The gender gap in political knowledge in Europe." *Social Politics: International Studies in Gender, State & Society* 21 (2): 261–289.

Fraile, Marta. 2014b. "Does deliberation contribute to decreasing the gender gap in knowledge?" *European Union Politics* 15 (3): 372–388.

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 Review* 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.

Graber, Doris A. 2001. *Processing politics: Learning from television in the Internet age*. University of Chicago Press.

Höhne, Jan Karem, Carina Cornesse, Stephan Schlosser, Mick P Couper, and Annelies G Blom. 2021. "Looking Up Answers to Political Knowledge Questions in Web Surveys." *Public Opinion Quarterly*.

Jerit, Jennifer, and Jason Barabas. 2017. "Revisiting the Gender Gap in Political Knowledge." *Political Behavior* 39 (4): 817–838.

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.

Lizotte, Mary-Kate, and Andrew H Sidman. 2009. "Explaining the gender gap in political knowledge." *Politics & Gender* 5 (02): 127–151.

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.

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

Luskin, Robert C, and John G Bullock. 2011. ““Don’t know” means “don’t know”: DK

responses and the public's level of political knowledge." *The Journal of Politics* 73 (02): 547–557.

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

McAllister, Ian. 2019. "The gender gap in political knowledge revisited: Australia's Julia Gillard as a natural experiment." *European Journal of Politics and Gender*.

McGlone, Matthew S, Joshua Aronson, and Diane Kobrynowicz. 2006. "Stereotype threat and the gender gap in political knowledge." *Psychology of Women Quarterly* 30 (4): 392–398.

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.

Pennebaker, James W, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of LIWC2015. Technical report.

Pereira, Frederico Batista. 2019. "Gendered Political Contexts: The Gender Gap in Political Knowledge." *The Journal of Politics* 81 (4): 1480–1493.

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.

Ryan, John Barry. 2011. "Accuracy and bias in perceptions of political knowledge." *Political Behavior* 33 (2): 335–356.

Spirling, Arthur. 2016. "Democratization and Linguistic Complexity The Effect of Franchise Extension on Parliamentary Discourse, 1832–1915." *The Journal of Politics* 78 (1): 120–136.

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.

Style, Hillary, and Jennifer Jerit. 2021. "Does it Matter if Respondents Look up Answers to Political Knowledge Questions?" *Public Opinion Quarterly*.

Tausczik, Yla R, and James W Pennebaker. 2010. "The psychological meaning of words: LIWC and computerized text analysis methods." *Journal of language and social psychology* 29 (1): 24–54.

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

Tetlock, Philip E. 1993. "Cognitive structural analysis of political rhetoric: Methodological and theoretical issues." In *Explorations in political psychology*. Duke University Press Durham.

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. 2020. "Self-Confidence and gender gaps in political interest, attention, and efficacy." *The Journal of Politics* 82 (4): 1490–1501.

Wolak, Jennifer, and Eric Gonzalez Juenke. 2019. "Descriptive representation and political knowledge." *Politics, Groups, and Identities*: 1–22.

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. 1990. "Political awareness, elite opinion leadership, and the mass survey response." *Social Cognition* 8 (1): 125.

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.

Appendices

Women Also Know Stuff: Challenging the Gender Gap in Political Sophistication

A Overview of Data Sources and Variables	2
I Summary of Surveys and Open-Ended Items	2
II Remaining Variables and Recoding	4
B Information on Discursive Sophistication Components	6
I Distribution of Word Counts & Proportion of Non-Response	6
II Pre-Processing and Topic Model Specification	8
III Discursive Sophistication Components	9
C Robustness Checks	10
I PreText Analysis	10
II Discursive Sophistication for Varying Model Specifications	11
III Controlling for Extraversion and Verbal Skills	14
D Tables of Model Estimates	16
I Figure 3: Effects of sophistication on turnout, political interest, internal efficacy, and external efficacy	16
II Figure 4: Expected information retrieval in the 2015 YouGov Study as a function of political sophistication	20
III Figure 7: The gender gap in political sophistication after controlling for common determinants.	21

Appendix A Overview of Data Sources and Variables

Throughout the paper, I apply the theoretical framework across multiple surveys fielded in two countries that employed different sets of open-ended questions. Each survey focuses on sophistication in the context of distinct political tasks, namely the evaluation of (1) candidates running for public office, (2) broad issue areas such as health care and gun legislation, and (3) specific legislative policy referenda. The data sets are briefly described below.

I Summary of Surveys and Open-Ended Items

2018 CES

- 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 care**, 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.

2012 & 2016 American National Election Study The main analyses are based on the 2012 and 2016 wave of the American National Election Study (ANES), each of which consists of a representative survey of about 5000 adults in the months before the US Presidential election in each year. About 2000 respondents in each wave participated in face-to-face interviews while the remaining respondents filled out the survey online. For the purpose of the present analyses, I rely on pooled datasets while controlling for differences in survey mode. Discursive sophistication is evaluated using a set of 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. 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. Individuals who did not respond to all of the open-ended items (420 in 2012; 204 in 2016), or who responded in Spanish (228 in 2012; 43 in 2016), are excluded from the analysis since discursive sophistication for them is not directly comparable to the remaining respondents.

2015 YouGov Survey In order to replicate and extend the main analyses, I rely on a separate nationally representative survey employing an alternative set of open-ended responses. The data was collected by YouGov in December 2015 and contains responses of 1000 U.S. citizens. 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.
- 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?

Compared to the open-ended likes/dislikes items included in the 2012 and 2016 ANES, the questions directly address considerations related to specific policy issues that were prominent in the political discourse at the time of the survey. Respondents who did not provide an answer to all of the open-ended questions were removed from the analysis (48).

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 \(2016\)](#). 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. However, respondents were only asked to justify their decision for or against a given proposition in verbatim if they participated in the vote in the first place. As such, about 5,000 individuals in the data set did not provide an open-ended response. The remaining respondents were asked to describe the main reason as well as additional justifications for their decision in two separate items. As before, discursive sophistication is evaluated based on verbatim responses to both questions. Since the Swiss survey was conducted in three different languages (German, French, and Italian), I computed the measure of discursive sophistication separately for each group of respondents.

II Remaining Variables and Recoding

Conventional measures of political knowledge:

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

Dependent variables:

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

Control variables:

- *Gender* (2012 & 2016 ANES, 2015 YouGov Survey): Dichotomous indicator for female respondents.
- *College education* (2012 & 2016 ANES, 2015 YouGov Survey): Dichotomous indicator for Bachelor’s degree or higher.

- *Family/Household income* (2012 & 2016 ANES, 2015 YouGov Survey): Self-reported household income categories.
- *Age* (2012 & 2016 ANES, 2015 YouGov Survey): Logged age in years.
- *Race* (2012 & 2016 ANES, 2015 YouGov Survey): Dichotomous indicator for black non-Hispanic vs. others.
- *Church attendance* (2012 & 2016 ANES, 2015 YouGov Survey): Six-category indicator of the frequency of church attendance.
- *Wordsum vocabulary scores* (2012 & 2016 ANES): Modified version of the GSS wordsum vocabulary test consisting of 10 terms.
- *Survey Mode* (2012 & 2016 ANES): Dichotomous indicator for face-to-face vs. online samples of the ANES surveys.
- *Extraversion* (2012 & 2016 ANES): Part of the Ten Item Personality Inventory (TIPI) measuring the “Big Five” personality traits.
- *Response length* (2012 & 2016 ANES): Logged number of words in the collection of open-ended responses by each individual.

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

I Distribution of Word Counts & Proportion of Non-Response

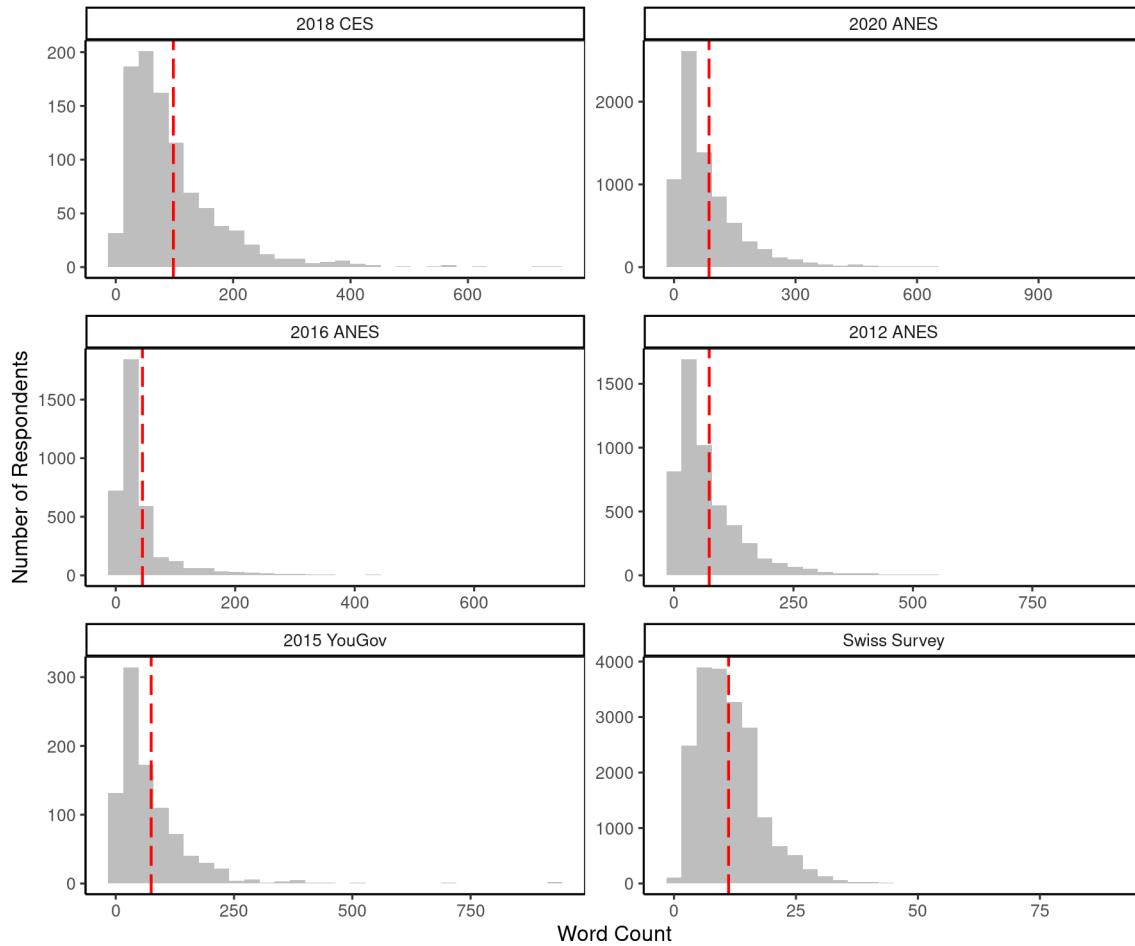


Figure B.1: Total word count across all open-ended responses for each survey participant. The dashed red lines indicate the average response lengths in each survey.

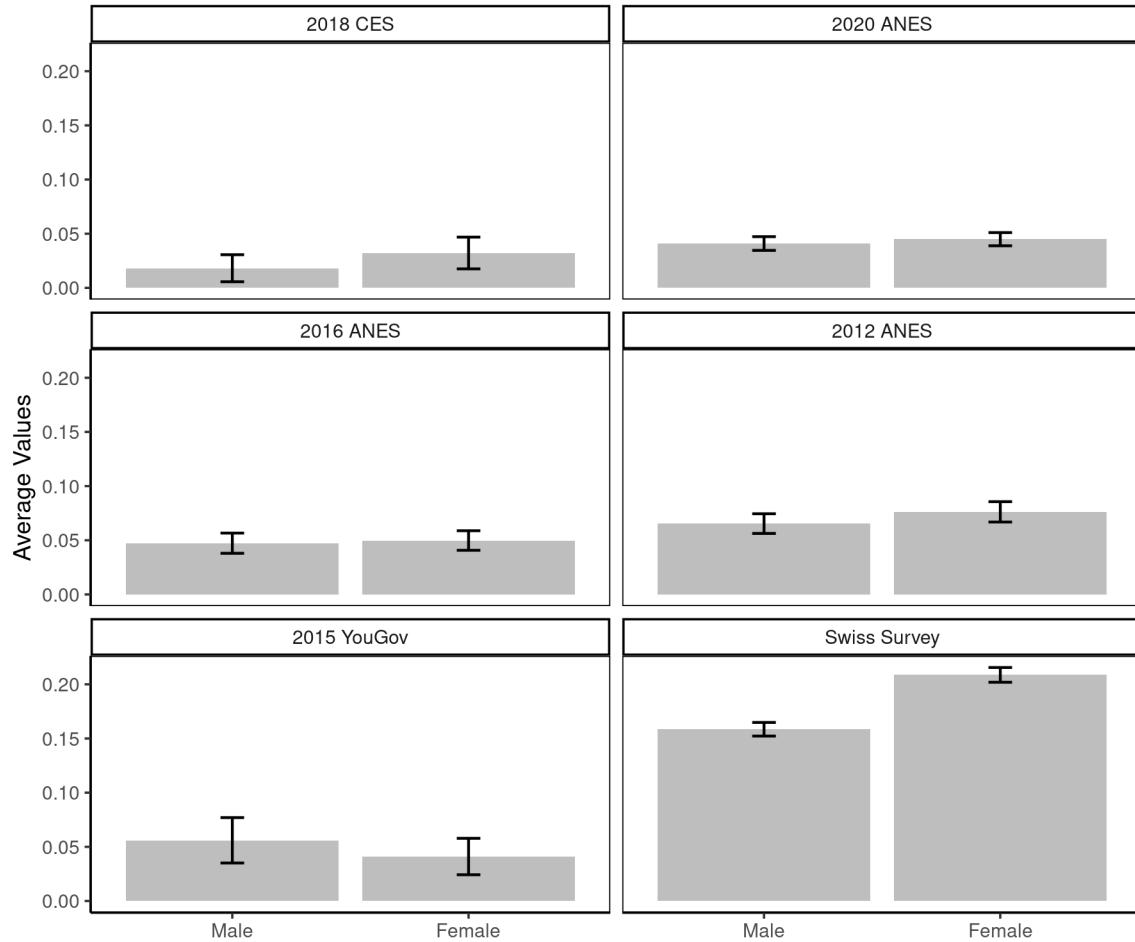


Figure B.2: Proportion of non-response comparing male and female survey participants (including 95% confidence intervals). Gender differences are only significant ($p < .05$) for the Swiss survey. The Swiss survey, however, contains an additional selection mechanism since respondents were only asked the set of open-ended questions if they previously reported that they voted in the respective referendum.

II Pre-Processing and Topic Model Specification

I rely on the structural topic model framework to extract and estimate differences in the topics mentioned by male and female respondents. I follow the guidelines in [Roberts et al. \(2014\)](#) to preprocess our open-ended responses (lowercasing, removing stopwords, removing punctuation, stemming) 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 including gender, this variable selection is equivalent to the procedure described in [Roberts et al. \(2014\)](#) for open-ended survey responses. In order to differentiate only broader considerations, we set the number of topics to be equal to 25, but we replicate equivalent results with larger numbers of topics below. Figure B.3 displays the resulting topic proportions along with their highest probability terms across all data sets included in the analyses.

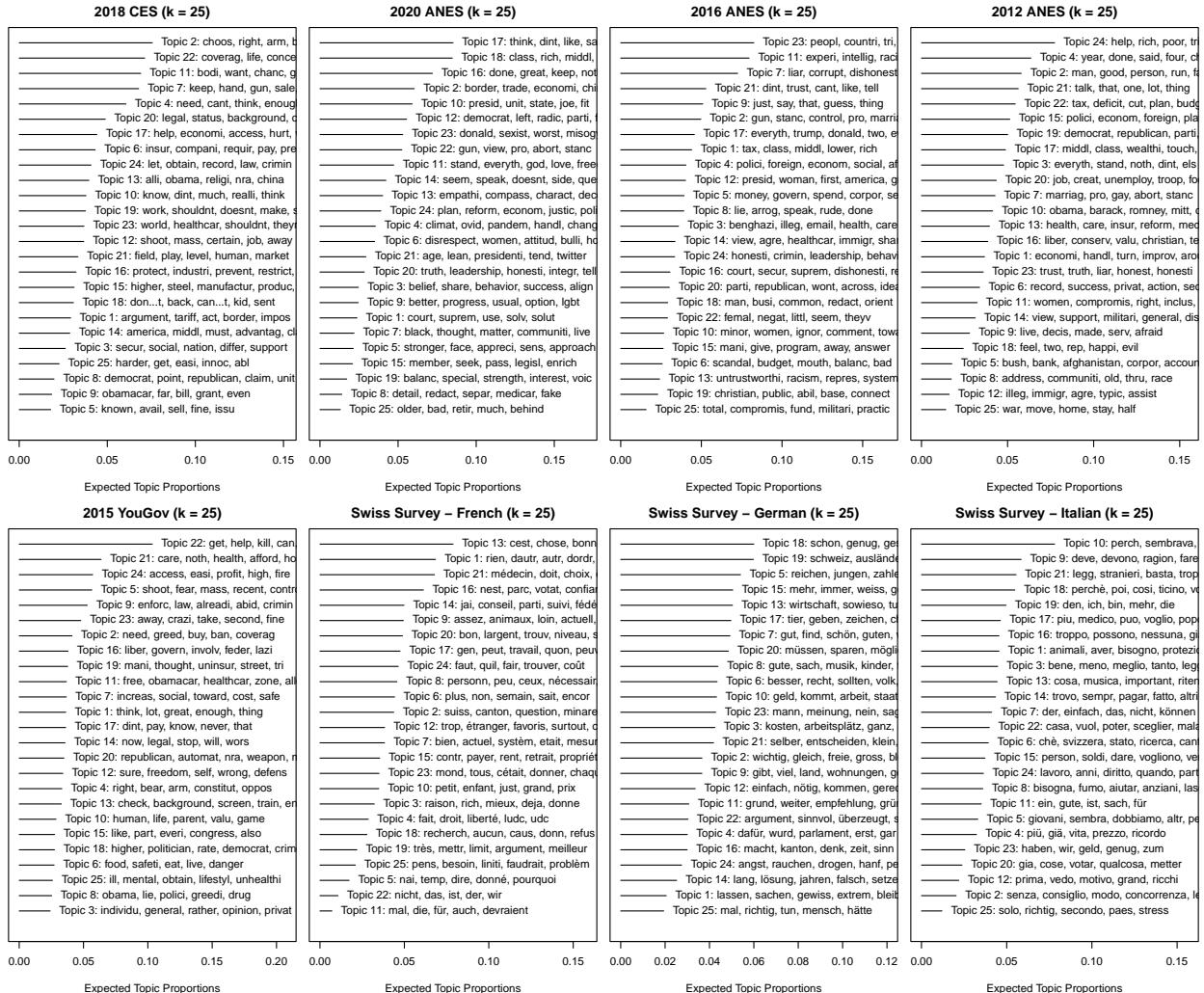


Figure B.3: Estimated topic proportions based on the structural topic model. See Appendix I for details on the model specification.

III Discursive Sophistication Components

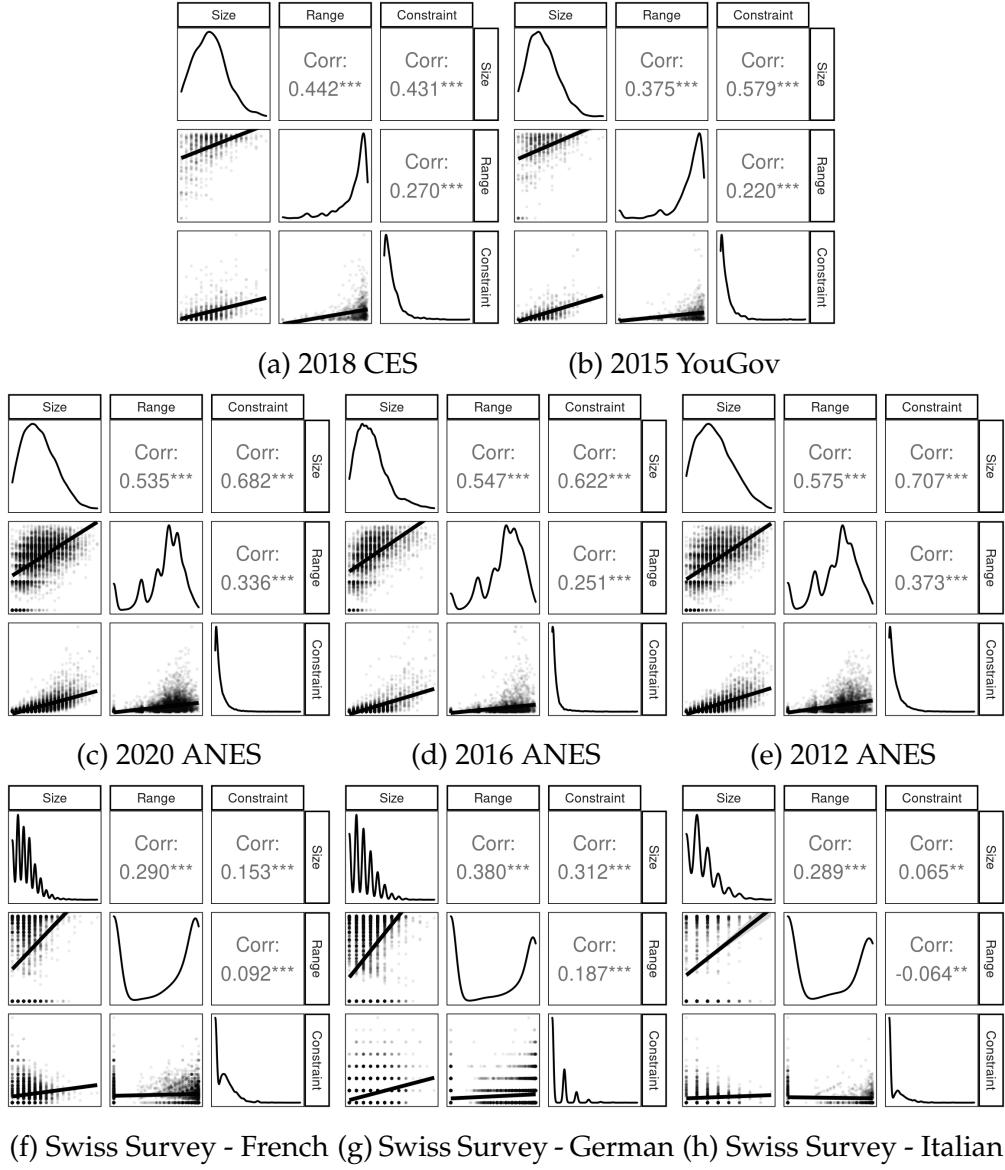


Figure B.4: Correlation matrix of individual components of discursive sophistication. The plots on the diagonal display univariate densities for each component. The panels in the lower triangular display the scatter plot of two measures as well as a linear fit.

Appendix C Robustness Checks

I PreText Analysis

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

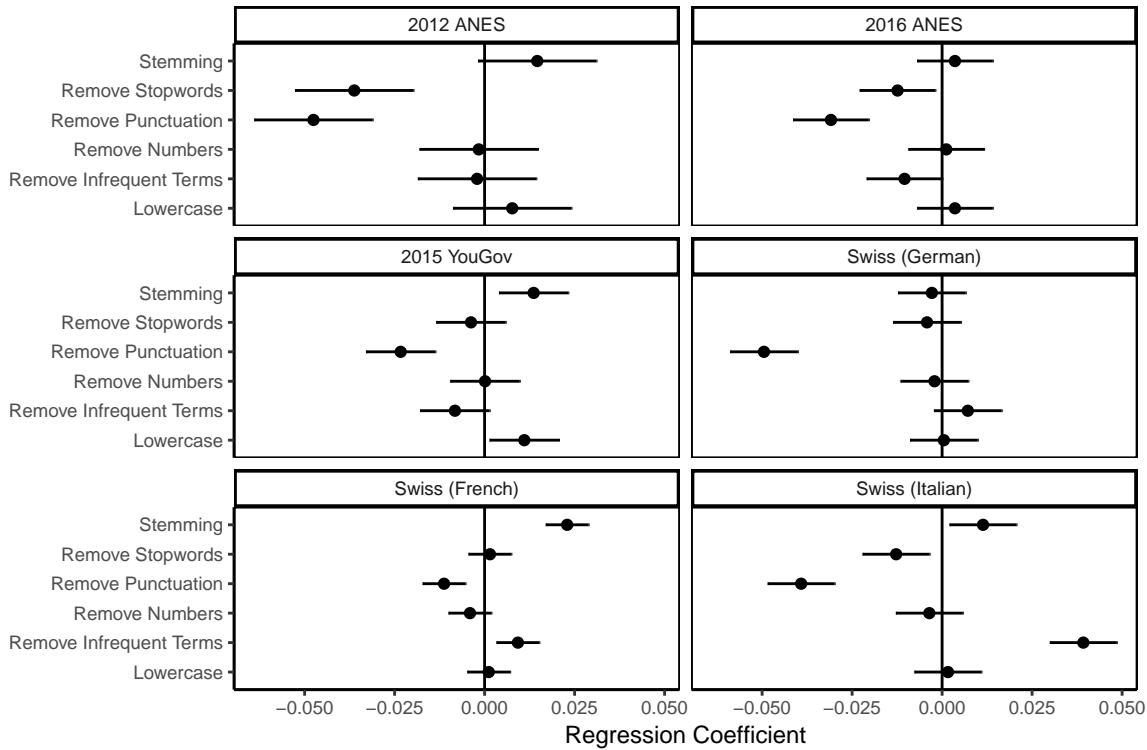


Figure C.1: PreText analysis of pre-processing decisions of open-ended responses across all datasets. Regression coefficients display the effects of each of the six pre-processing choices on the resulting preText score.

II Discursive Sophistication for Varying Model Specifications

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

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

Figure C.2 examines whether the proposed measure of discursive sophistication is sensitive to the removal of infrequent terms as well as the chosen number of topics k . The y-axis depicts the preferred pre-processing regime including all steps discussed above while the x-axis plots results for alternative specifications. The panels on the left compare the preferred specification to discursive sophistication based on a reduced number of topics ($k = 20$). The middle panels additionally include infrequent terms instead of removing them.¹⁵ The panels on the right do not perform stemming as part of the pre-processing step. Across all panels, discursive sophistication scores are highly correlated and therefore insensitive to pre-processing choices and varying numbers of topics.

In summary, open-ended responses in the analyses reported in the main text are pre-processed by stemming and lowercasing, as well as the removing numbers, punctuation, stopwords, and infrequent terms (i.e., terms that appear in fewer than 10 responses).¹⁶ While the results discussed in the manuscript are based on this preferred specification, the substantive results are robust for alternative pre-processing regimes or varying numbers

¹⁵Calculating discursive sophistication with large numbers of topics while including infrequent terms is computationally prohibitive.

¹⁶Prior to applying these pre-processing steps, open-ended responses in the 2012 & 2016 ANES as well as the 2015 YouGov survey are cleaned by correcting spelling errors using an implementation of the Aspell spell-checking algorithm (www.aspell.net).

of topics.

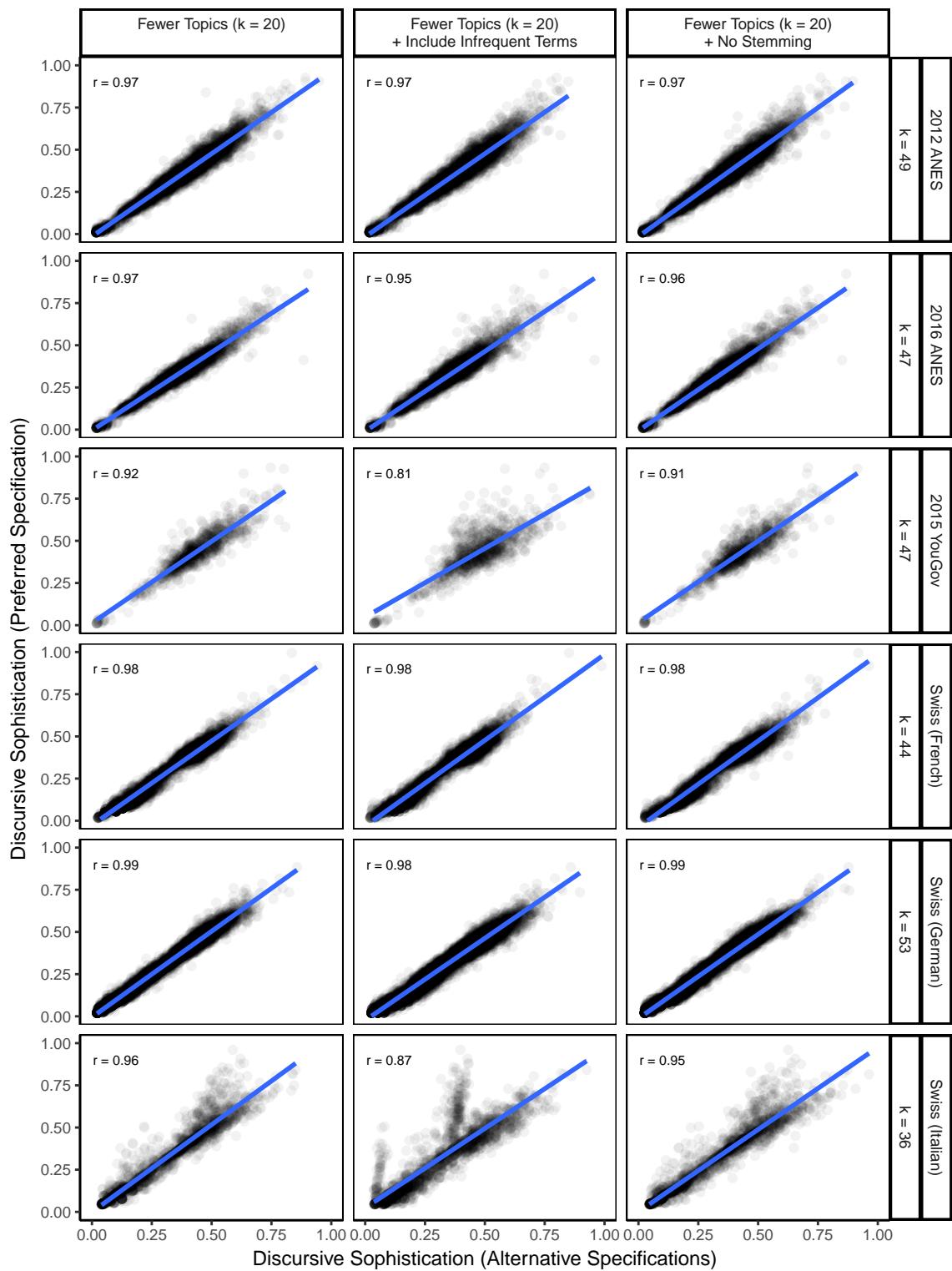


Figure C.2: Robustness of discursive sophistication measure for different pre-processing choices and topic model specifications.

III Controlling for Extraversion and Verbal Skills

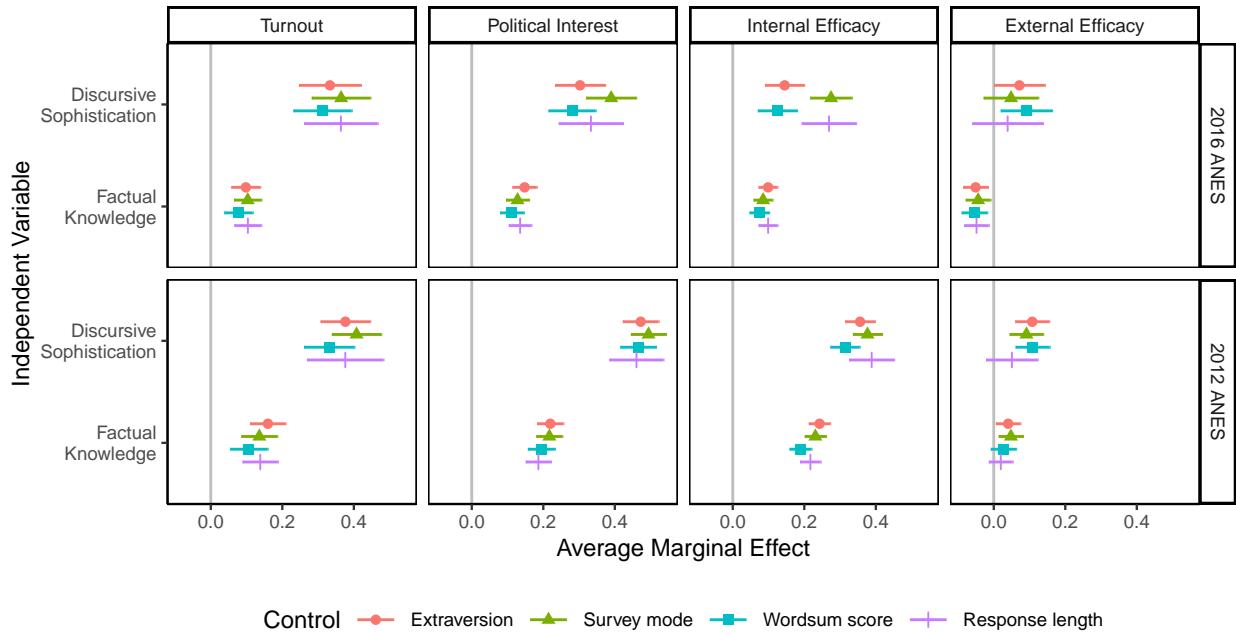


Figure C.3: Effects of sophistication on internal efficacy, external efficacy, non-conventional participation, and turnout in the 2012 and 2016 ANES. For each dependent variable, the figure displays the change in expected values after increasing each sophistication measure from -1 to +1 standard deviation from its mean (including 95% confidence intervals). Model estimates are based on logistic regression (turnout) or OLS (internal efficacy, external efficacy, non-conventional participation). Both sophistication measure are included simultaneously while controlling for gender, education, income, age, race, church attendance, survey mode, Wordsum vocabulary scores, as well as personality characteristics (extraversion and being reserved).

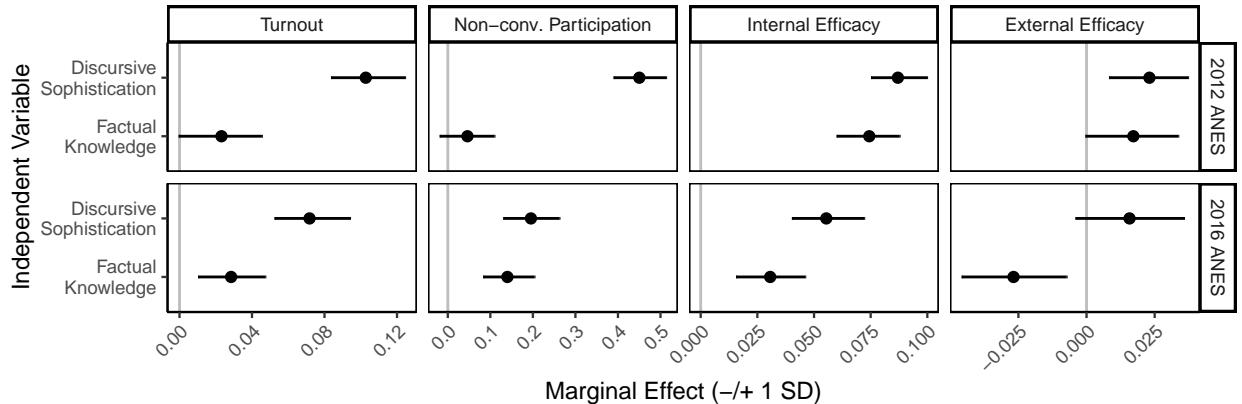


Figure C.4: Effects of sophistication on internal efficacy, external efficacy, non-conventional participation, and turnout in the 2012 and 2016 ANES. For each dependent variable, the figure displays the change in expected values after increasing each sophistication measure from -1 to +1 standard deviation from its mean (including 95% confidence intervals). Model estimates are based on logistic regression (turnout) or OLS (internal efficacy, external efficacy, non-conventional participation). Both sophistication measure are included simultaneously while controlling for gender, education, income, age, race, church attendance, survey mode, Wordsum vocabulary scores, as well as personality characteristics (extraversion and being reserved).

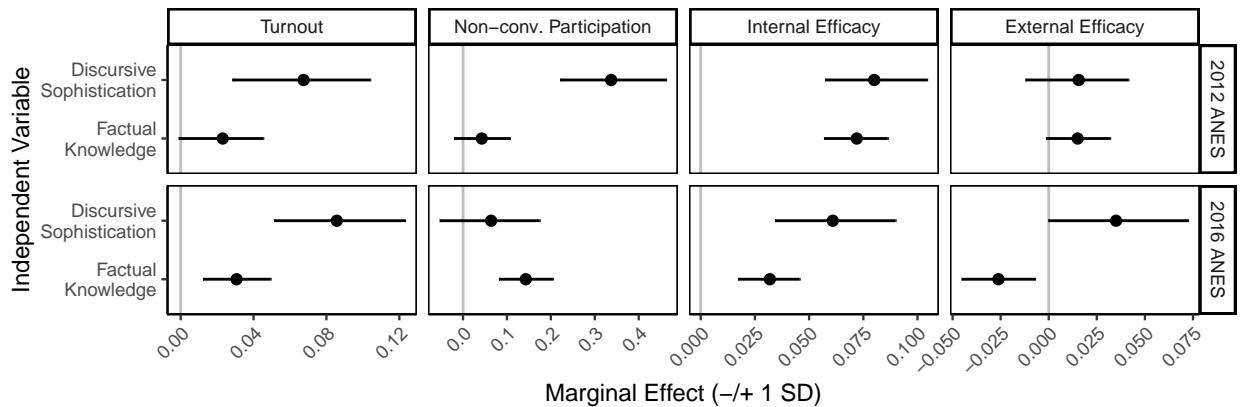


Figure C.5: Effects of sophistication on internal efficacy, external efficacy, non-conventional participation, and turnout in the 2012 and 2016 ANES. For each dependent variable, the figure displays the change in expected values after increasing each sophistication measure from -1 to +1 standard deviation from its mean (including 95% confidence intervals). Model estimates are based on logistic regression (turnout) or OLS (internal efficacy, external efficacy, non-conventional participation). Both sophistication measure are included simultaneously while controlling for gender, education, income, age, race, church attendance, survey mode, Wordsum vocabulary scores, as well as the logged word count across open-ended responses.

Appendix D Tables of Model Estimates

I Figure 3: Effects of sophistication on turnout, political interest, internal efficacy, and external efficacy

Table D.1: Effects of sophistication on turnout, political interest, internal efficacy, and external efficacy in the 2018 CES. Standard errors in parentheses. Estimates are used for Figure 3 in the main text.

	Dependent variable:							
	Turnout		Political Interest		Internal Efficacy		External Efficacy	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Discursive Soph.	2.793*** (0.738)		0.400*** (0.073)		0.445*** (0.066)		-0.046 (0.075)	
Factual Knowledge		2.189*** (0.258)		0.303*** (0.025)		0.241*** (0.024)		0.002 (0.028)
Female	-0.484** (0.184)	-0.313 (0.195)	-0.086*** (0.017)	-0.062*** (0.017)	-0.062*** (0.016)	-0.043** (0.016)	0.016 (0.018)	0.016 (0.018)
College Degree	0.758*** (0.200)	0.558** (0.211)	0.058** (0.019)	0.034 (0.018)	0.068*** (0.017)	0.056*** (0.016)	0.041* (0.019)	0.038* (0.019)
Household Income	1.084* (0.446)	0.682 (0.476)	0.180*** (0.042)	0.129** (0.040)	0.125** (0.039)	0.087* (0.038)	0.034 (0.044)	0.032 (0.044)
Age	0.045*** (0.006)	0.039*** (0.006)	0.004*** (0.0005)	0.003*** (0.0005)	0.002*** (0.0004)	0.002*** (0.0004)	-0.001 (0.001)	-0.001 (0.001)
Black	-0.670* (0.262)	-0.801** (0.273)	-0.055 (0.029)	-0.061* (0.027)	0.013 (0.026)	0.004 (0.025)	-0.026 (0.030)	-0.024 (0.029)
Church Attendance	0.598* (0.274)	0.700* (0.290)	0.035 (0.025)	0.040 (0.024)	-0.012 (0.023)	-0.006 (0.022)	0.085** (0.026)	0.084** (0.026)
Constant	-2.541*** (0.435)	-2.557*** (0.375)	0.215*** (0.042)	0.226*** (0.033)	0.257*** (0.039)	0.313*** (0.031)	0.325*** (0.044)	0.307*** (0.036)
Observations	868	868	867	867	866	866	866	866
R ²			0.225	0.311	0.179	0.227	0.021	0.021
Log Likelihood	-403.885	-372.851						

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 2020 ANES. Standard errors in parentheses. Estimates are used for Figure 3 in the main text.

	Dependent variable:							
	Turnout		Political Interest		Internal Efficacy		External Efficacy	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Discursive Soph.	3.263*** (0.313)		0.450*** (0.022)		0.094*** (0.017)		0.067** (0.026)	
Factual Knowledge		1.167*** (0.155)		0.174*** (0.011)		0.048*** (0.008)		0.070*** (0.012)
Female	0.365*** (0.084)	0.416*** (0.084)	-0.056*** (0.006)	-0.041*** (0.006)	-0.023*** (0.004)	-0.018*** (0.004)	0.021** (0.007)	0.028*** (0.007)
College Degree	0.803*** (0.098)	0.854*** (0.098)	0.030*** (0.006)	0.033*** (0.006)	-0.010* (0.005)	-0.010* (0.005)	0.074*** (0.007)	0.068*** (0.007)
Household Income	1.356*** (0.141)	1.449*** (0.140)	0.030** (0.010)	0.036*** (0.010)	0.015* (0.007)	0.015* (0.007)	0.062*** (0.011)	0.058*** (0.011)
Age	0.032*** (0.003)	0.032*** (0.003)	0.004*** (0.0002)	0.004*** (0.0002)	0.0004*** (0.0001)	0.0003* (0.0001)	0.0002 (0.0002)	-0.00002 (0.0002)
Black	0.230 (0.137)	0.117 (0.134)	0.016 (0.010)	0.008 (0.010)	-0.012 (0.008)	-0.013 (0.008)	0.027* (0.012)	0.028* (0.012)
Church Attendance	0.558*** (0.143)	0.567*** (0.141)	-0.013 (0.009)	-0.010 (0.009)	-0.031*** (0.007)	-0.031*** (0.007)	0.013 (0.010)	0.014 (0.010)
Constant	-1.871*** (0.166)	-1.519*** (0.158)	0.329*** (0.012)	0.387*** (0.012)	0.568*** (0.010)	0.576*** (0.009)	0.237*** (0.015)	0.231*** (0.014)
Observations	6,370	6,370	6,964	6,964	6,317	6,317	6,315	6,315
R ²			0.165	0.149	0.018	0.019	0.038	0.042
Log Likelihood	-1,990.469		-2,017.154					

Note:

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

Table D.3: Effects of sophistication on turnout, political interest, internal efficacy, and external efficacy in the 2016 ANES. Standard errors in parentheses. Estimates are used for Figure 3 in the main text.

	Dependent variable:							
	Turnout		Political Interest		Internal Efficacy		External Efficacy	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Discursive Soph.	3.623*** (0.445)		0.305*** (0.033)		0.155*** (0.028)		0.080* (0.036)	
Factual Knowledge		1.069*** (0.205)		0.131*** (0.016)		0.098*** (0.013)		-0.050** (0.017)
Female	0.159 (0.111)	0.155 (0.110)	-0.067*** (0.008)	-0.064*** (0.008)	-0.057*** (0.007)	-0.055*** (0.007)	0.0004 (0.009)	-0.002 (0.009)
College Degree	0.720*** (0.136)	0.832*** (0.134)	0.067*** (0.009)	0.076*** (0.009)	0.075*** (0.008)	0.078*** (0.008)	0.056*** (0.010)	0.063*** (0.010)
Household Income	0.787*** (0.199)	0.859*** (0.197)	0.048** (0.015)	0.046** (0.015)	0.053*** (0.013)	0.049*** (0.013)	0.054** (0.017)	0.064*** (0.017)
Age	0.025*** (0.003)	0.023*** (0.003)	0.004*** (0.0002)	0.004*** (0.0002)	0.001*** (0.0002)	0.001*** (0.0002)	-0.0001 (0.0003)	0.00002 (0.0003)
Black	1.014*** (0.232)	0.934*** (0.229)	0.021 (0.015)	0.016 (0.015)	0.042*** (0.012)	0.039** (0.012)	-0.015 (0.016)	-0.017 (0.016)
Church Attendance	1.020*** (0.186)	1.079*** (0.186)	0.005 (0.012)	0.011 (0.012)	-0.012 (0.010)	-0.009 (0.010)	0.080*** (0.013)	0.080*** (0.013)
Constant	-1.248*** (0.216)	-0.734*** (0.200)	0.322*** (0.018)	0.352*** (0.017)	0.420*** (0.015)	0.426*** (0.014)	0.299*** (0.020)	0.339*** (0.019)
Observations	3,562	3,562	3,582	3,582	3,104	3,104	3,106	3,106
R ²			0.144	0.139	0.099	0.105	0.037	0.038
Log Likelihood	-1,129.717		-1,150.989					

Note:

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

Table D.4: Effects of sophistication on turnout, political interest, internal efficacy, and external efficacy in the 2012 ANES. Standard errors in parentheses. Estimates are used for Figure 3 in the main text.

	Dependent variable:							
	Turnout		Political Interest		Internal Efficacy		External Efficacy	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Discursive Soph.	2.978*** (0.289)		0.480*** (0.025)		0.357*** (0.021)		0.108*** (0.024)	
Factual Knowledge		1.249*** (0.198)		0.213*** (0.018)		0.234*** (0.015)		0.028 (0.017)
Female	0.025 (0.083)	0.062 (0.083)	-0.074*** (0.007)	-0.061*** (0.007)	-0.063*** (0.006)	-0.050*** (0.006)	0.016* (0.007)	0.017* (0.007)
College Degree	0.552*** (0.108)	0.625*** (0.107)	0.044*** (0.008)	0.055*** (0.008)	0.048*** (0.007)	0.049*** (0.007)	0.037*** (0.008)	0.041*** (0.008)
Household Income	1.216*** (0.150)	1.188*** (0.151)	0.022 (0.013)	0.019 (0.013)	0.048*** (0.011)	0.034** (0.011)	0.013 (0.012)	0.015 (0.012)
Age	0.029*** (0.002)	0.027*** (0.003)	0.003*** (0.0002)	0.003*** (0.0002)	0.001** (0.0002)	0.0001 (0.0002)	-0.0003 (0.0002)	-0.0003 (0.0002)
Black	0.714*** (0.118)	0.839*** (0.119)	0.027** (0.009)	0.047*** (0.010)	0.037*** (0.008)	0.059*** (0.008)	0.083*** (0.009)	0.086*** (0.009)
Church Attendance	0.637*** (0.126)	0.681*** (0.125)	0.008 (0.010)	0.015 (0.010)	0.003 (0.008)	0.009 (0.008)	0.049*** (0.010)	0.050*** (0.010)
Constant	-1.745*** (0.161)	-1.545*** (0.168)	0.291*** (0.015)	0.312*** (0.016)	0.413*** (0.012)	0.394*** (0.013)	0.310*** (0.014)	0.324*** (0.015)
Observations	4,714	4,714	5,002	5,002	4,994	4,994	4,983	4,983
R ²			0.164	0.126	0.126	0.117	0.038	0.034
Log Likelihood	-1,900.648		-1,935.980					

Note:

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

II Figure 4: Expected information retrieval in the 2015 YouGov Study as a function of political sophistication

Table D.5: Effects of sophistication on information retrieval in the 2015 YouGov study. Standard errors in parentheses. Estimates are used for Figure 4 in the main text.

	<i>Dependent variable:</i> Information Retrieval	
	(1)	(2)
Discursive Soph.	0.348*** (0.050)	
Factual Knowledge		0.170*** (0.026)
Female	0.022 (0.012)	0.052*** (0.012)
College Degree	0.031* (0.014)	0.019 (0.014)
Household Income	-0.030 (0.030)	-0.070* (0.030)
Age	0.001** (0.0004)	0.001* (0.0004)
Black	-0.035 (0.021)	-0.038 (0.021)
Church Attendance	-0.065*** (0.017)	-0.059*** (0.017)
Constant	0.651*** (0.028)	0.695*** (0.025)
Observations	792	792
R ²	0.110	0.104

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

III Figure 7: The gender gap in political sophistication after controlling for common determinants.

Table D.6: Effects of gender on discursive sophistication in the CES, ANES, and YouGov study. Estimates are used for Figure 7 in the main text.

	<i>Dependent variable:</i>				
	Discursive Sophistication				
	2018 CES	2020 ANES	2016 ANES	2012 ANES	2015 YouGov
	(1)	(2)	(3)	(4)	(5)
Female	-0.003 (0.008)	-0.007* (0.003)	-0.004 (0.004)	-0.00002 (0.004)	0.009 (0.008)
College Degree	0.049*** (0.009)	0.058*** (0.003)	0.053*** (0.005)	0.068*** (0.005)	0.019 (0.010)
Household Income	0.018 (0.020)	0.059*** (0.005)	0.048*** (0.008)	0.061*** (0.007)	-0.001 (0.021)
Age	0.001*** (0.0002)	0.001*** (0.0001)	0.0004** (0.0001)	0.001*** (0.0001)	0.001** (0.0003)
Black	-0.037** (0.013)	-0.037*** (0.005)	-0.019** (0.007)	0.003 (0.005)	-0.052*** (0.015)
Church Attendance	0.008 (0.012)	0.005 (0.005)	0.010 (0.006)	0.011 (0.006)	-0.007 (0.012)
Constant	0.374*** (0.015)	0.253*** (0.006)	0.259*** (0.008)	0.248*** (0.008)	0.364*** (0.015)
Observations	868	6,965	3,582	5,004	792
R ²	0.074	0.109	0.076	0.092	0.036

Note:

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

Table D.7: Effects of gender on factual knowledge in the CES, ANES, and YouGov study. Estimates are used for Figure 7 in the main text.

	<i>Dependent variable:</i>				
	Factual Knowledge				
	2018 CES	2020 ANES	2016 ANES	2012 ANES	2015 YouGov
	(1)	(2)	(3)	(4)	(5)
Female	-0.083*** (0.022)	-0.104*** (0.006)	-0.026** (0.009)	-0.059*** (0.006)	-0.158*** (0.016)
College Degree	0.143*** (0.023)	0.129*** (0.007)	0.057*** (0.009)	0.101*** (0.006)	0.111*** (0.019)
Household Income	0.191*** (0.053)	0.117*** (0.011)	0.120*** (0.016)	0.153*** (0.010)	0.233*** (0.041)
Age	0.005*** (0.001)	0.004*** (0.0002)	0.002*** (0.0002)	0.003*** (0.0002)	0.003*** (0.0005)
Black	-0.031 (0.036)	-0.051*** (0.011)	-0.010 (0.015)	-0.088*** (0.008)	-0.085** (0.029)
Church Attendance	-0.007 (0.032)	-0.006 (0.010)	-0.023 (0.013)	-0.010 (0.008)	-0.051* (0.023)
Constant	0.458*** (0.041)	0.320*** (0.013)	0.372*** (0.017)	0.460*** (0.011)	0.486*** (0.029)
Observations	868	6,965	3,582	5,004	792
R ²	0.183	0.172	0.067	0.239	0.275

Note:

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