

Women Also Know Stuff

Challenging the Gender Gap in Political Sophistication*

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 $N \approx 35,000$). 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://womentalsoknowstuff.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.

Political sophistication is one of the most fundamental concepts in the study of political attitudes and behavior—a crucial moderator impacting a range of mechanisms such as political decision-making and vote choice (Macdonald, Rabinowitz, and Listhaug, 1995; Lau and Redlawsk, 2001), persuasion and motivated reasoning (Zaller, 1992; Lodge and Taber, 2013), or the susceptibility to misinformation (Vegetti and Mancosu, 2020). Yet, fundamental issues regarding the measurement of political sophistication 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 (Delli 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, 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 may reflect genuine differences in political interest between men and women due to gendered socialization (Bos et al., 2021). At least part of the observed gender gap, however, can be attributed to how we measure political knowledge. For instance, men are more willing to guess when answering recall questions, which can inflate their estimated knowledge levels (Mondak and Anderson, 2004; Fortin-Rittberger, 2020). Other research finds 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 more equitable representation of women (Pereira, 2019; Kraft and Dolan, 2022).

In this paper, I re-examine the gender gap by proposing a measure of *discursive sophistication* 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 $N \approx 35,000$) 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

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

sum, the results suggest that exploring open-ended responses provides new opportunities to examine political sophistication across time and contexts.

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—[Delli Carpini and Keeter \(1993\)](#)—has been cited more than 1000 times since its publication. In other words, political knowledge remains a concept of intense scholarly interest—and that it is frequently measured using standard off-the-shelf recall questions.

The ubiquity of basic recall questions in public opinion research is accompanied by the frequent findings that many people know too little about politics ([Delli Carpini and Keeter, 1996](#); [Barabas et al., 2014](#)) and that 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 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 exclusive focus 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

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?”

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 revolves 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 instance, 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 con-

text 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, 2021). 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 relevance for competent decision-making in politics, which lead some scholars to suggest that we should start considering alternatives to these types of recall questions (Druckman, 2014). From a theoretical perspective, knowledge scores are 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).

Political sophistication can then be characterized by how these ideas and attitudes (or considerations) are structured 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 my approach in the context of a concrete example, consider a questionnaire that asks respondents to answer the following open-ended item:

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.

Now suppose that this questionnaire includes a whole set of similar prompts on other topics such as **abortion**, **immigration**, **health care**, and **trade policies**—each asking re-

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.

spondents 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 maximum number of topics observed across individuals.

Second, the *range* of a belief system is defined as the dispersion of considerations

⁴Please refer to the appendix for additional information. Specifically, see Appendix A for a data overview and Appendix B for descriptive information on open-ended responses, pre-processing, and modeling choices for the structural topic models. Appendix C contains additional robustness checks including a preText analysis to explore sensitivity for alternative model specifications (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).

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 indi-

vidual 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.

Data and Analytical Strategy

To evaluate my proposed measure of discursive sophistication, I included the battery of open-ended questions described above in a 2018 wave of the Cooperative Election Study (CES),⁷ which consists of a national stratified sample of 1,000 respondents. In addition, I illustrate the versatility and robustness of the approach by applying the measure across multiple previously collected surveys that employ a range of alternative open-ended items. Below is a summary of all data sets and items used in the subsequent analysis:⁸

- **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

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.

⁷Formerly Cooperative Congressional Election Study (CCES).

⁸A detailed description of each data set and the specific question wording is included in Appendix A.

different subsets of the studies listed above, depending on the availability of necessary 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.⁹ 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 (see Table 1).¹⁰

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, 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 presidential term limits or the majority party in either chamber of Congress. 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).¹¹

Figure 1 compares discursive sophistication to conventional knowledge metrics for the CES and ANES. Each figure presents scatterplots between individual measures (lower tri-

⁹See Appendix B.III for correlation matrices between individual components.

¹⁰I rely on the CES and ANES here since these surveys employ a larger set of open-ended questions.

¹¹Interviewer assessments were only recorded in the face-to-face sample of the 2012 and 2016 ANES.

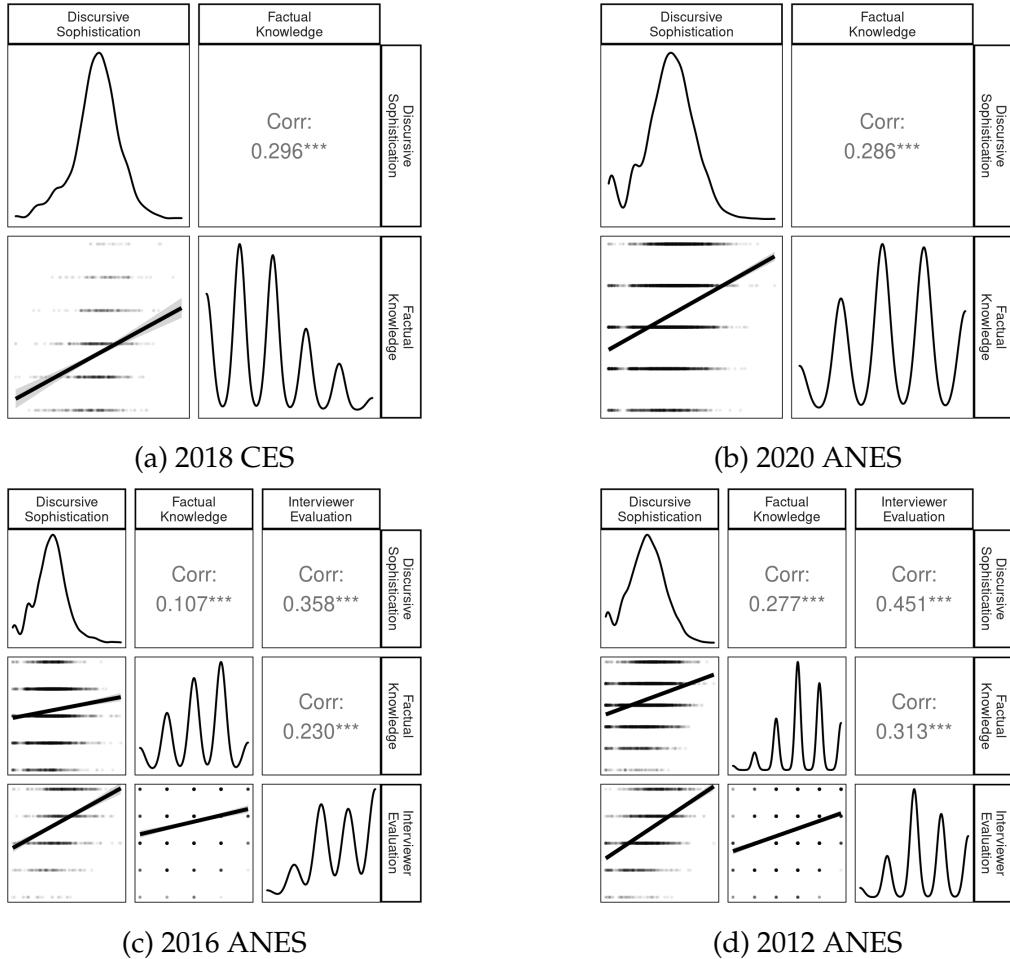


Figure 1: Correlation matrix of discursive sophistication and conventional political knowledge metrics. The plots on the diagonal display univariate densities for each variable. The panels in the lower triangular display the scatter plot of two measures as well as a linear fit. The upper triangular displays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

angular), 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. In other words, 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 from 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. Note that the entries are original responses without editing.

The results are presented in Table 2. Each row represents one of the open-ended responses 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 2 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.¹²

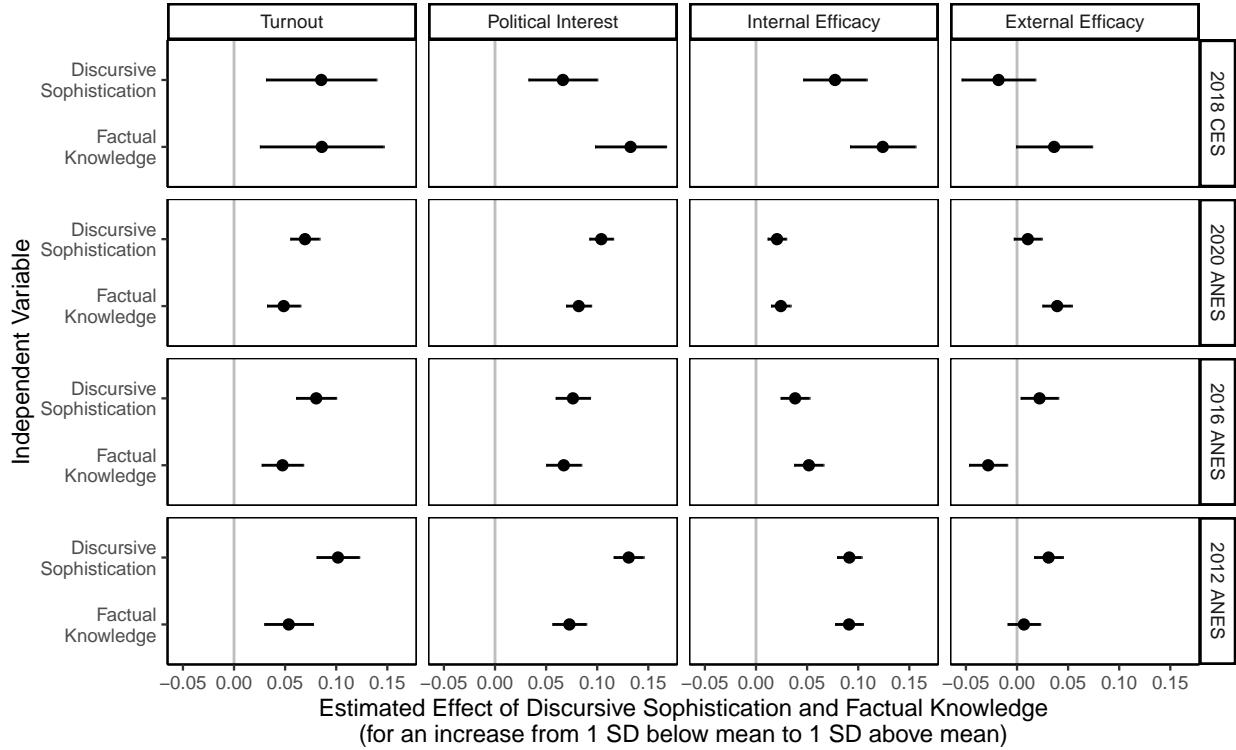


Figure 2: Effects of sophistication on turnout, political interest, internal efficacy, and external efficacy in the CES and ANES. For each dependent variable, the figure displays the average marginal effects (AME) for each sophistication measure (including 95% confidence intervals). Model estimates are based on logistic regression (turnout) or OLS (political interest, internal efficacy, external efficacy). Each analysis includes controls for sociodemographic variables. Full regression results are displayed in the appendix, Tables D.1 through D.4.

Each panel in Figure 2 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 are more ambiguous. Factual knowledge has strikingly inconsistent effects—sometimes predicting higher, lower, or no change in external efficacy. Discursive sophis-

¹²See Appendix D for full regression results.

tication, 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. As a robustness check, Appendix C.III provides additional regression results controlling for various 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. residents. 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 make it easier for people to incorporate relevant new information about parties, office-holders, and policies. After all, Zaller (1990, 1992) and others argue that factual knowledge is the best available proxy for political 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 3 displays the relationship between political sophistication and disease infor-

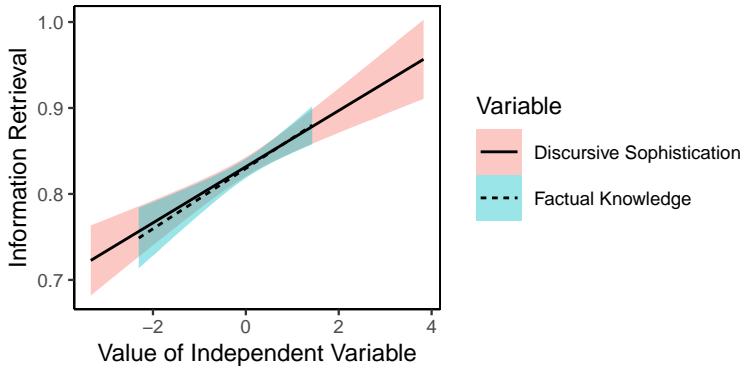


Figure 3: Expected information retrieval in the 2015 YouGov Study as a function of political sophistication (including 95% confidence intervals). Estimates are based on a linear regression including controls for sociodemographic variables. Full regression results are displayed in the appendix, Table D.5.

mation retrieval in the 2015 YouGov study. Estimates are based on linear regression 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 are positively correlated with 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 manually coded open-ended responses across three languages. Colombo (2018) 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 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 (2018) 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 (2018) manual coding with my automated measure corroborates the face validity of discursive sophistication.

The results are presented in Figure 4, which displays the distribution of discursive sophistication for each level of justification coded by Colombo (2018) 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). 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 new information from news sources, and shares significant overlap with manually coded levels of justification in open-ended responses. Next, I illustrate how discursive sophistication can help refine previous findings regarding the gender gap in political knowledge.

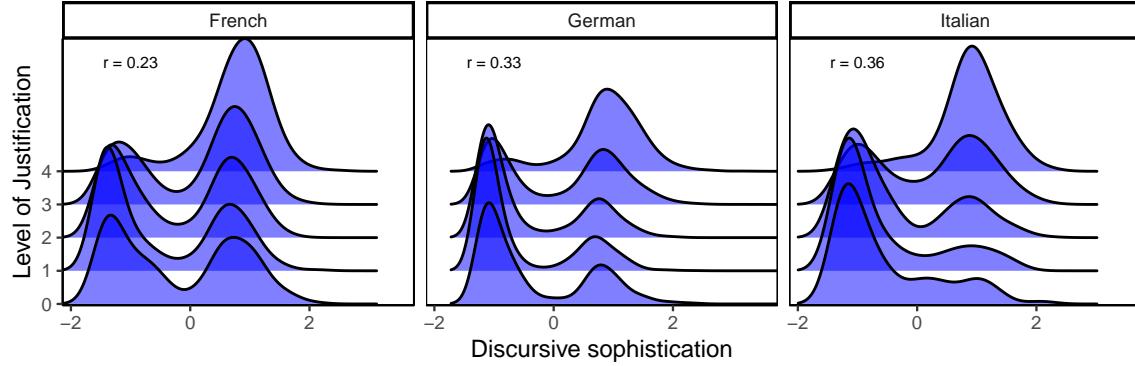


Figure 4: Discursive sophistication and manually coded level of justification (Colombo, 2018) in Swiss post-referendum surveys. The plot compares kernel densities of discursive sophistication for each manually coded level of justification.

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 5 displays the average levels of discursive sophistication and conventional metrics comparing both genders. While we observe sizable and statistically significant gender gaps in factual knowledge across the CES, ANES, and YouGov surveys, these differences disappear for discursive sophistication. Even though women do not perform as well as men on political quizzes, they do not differ substantially in the level of elaboration when describing their political preferences.

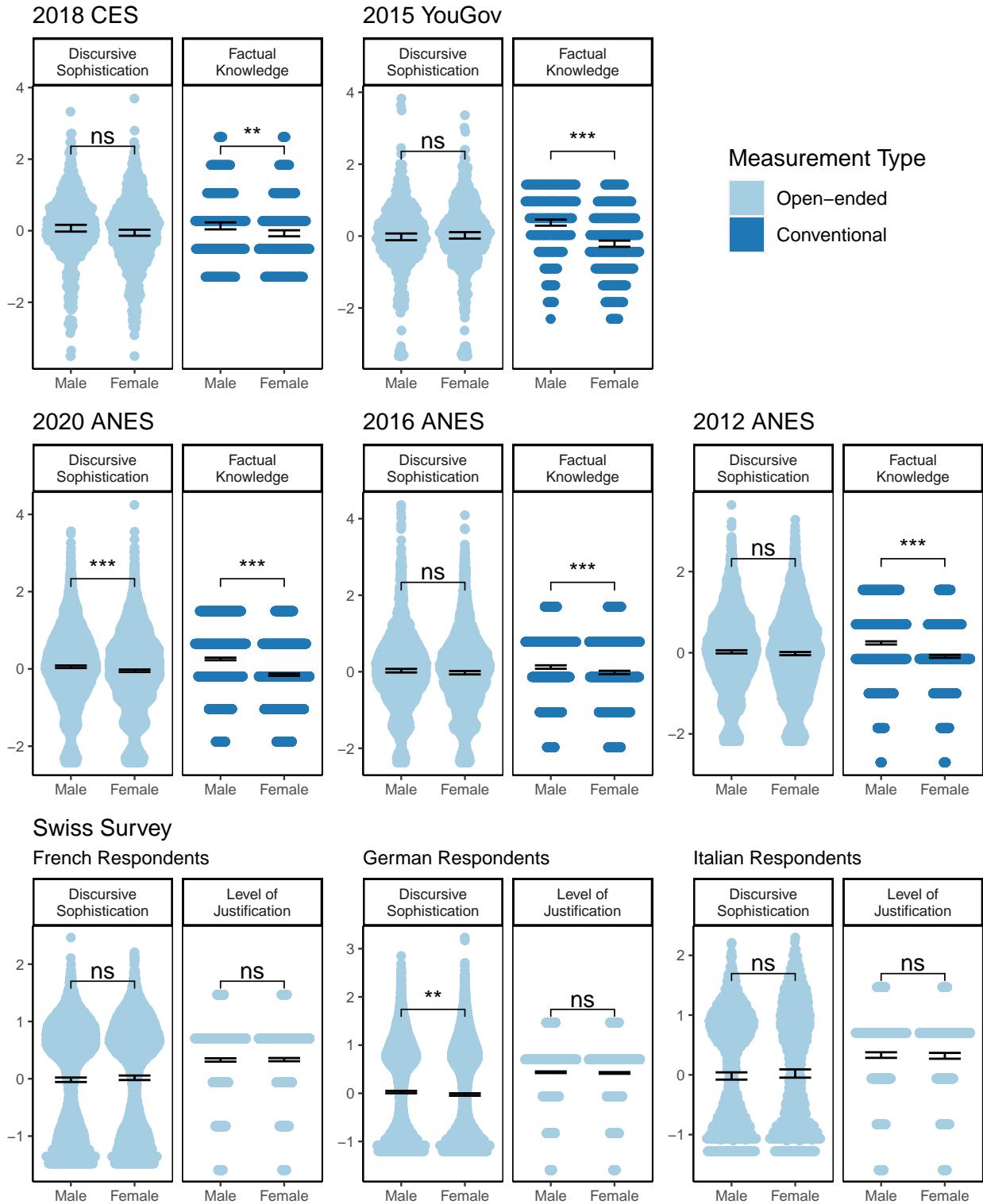


Figure 5: The gender gap in political sophistication. The figures display mean levels of political sophistication using open-ended (light blue) or conventional (dark blue) measures comparing women and men (including 95% confidence intervals). All gender differences in factual knowledge are statistically 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 (2018) 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 5 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 6 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.¹³ The gender gap in factual political knowledge, however, persists and is sub-

¹³I 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.

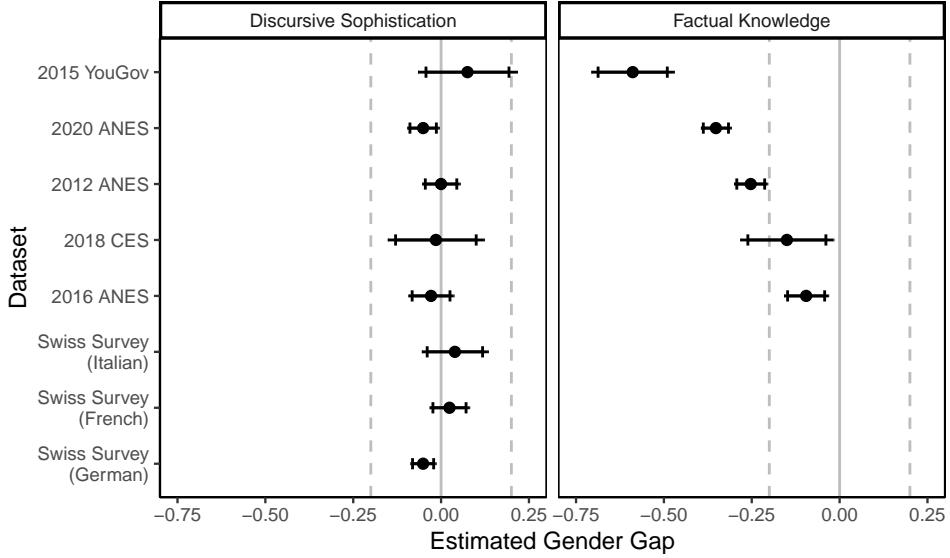


Figure 6: The gender gap in political sophistication controlling for common determinants. Estimates are OLS regression coefficients with 95% confidence intervals. Dependent variables are discursive sophistication and factual political knowledge. Estimates are based on a linear regression including controls for sociodemographic variables. Full regression 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.¹⁴ 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.

¹⁴See Appendix D for full regression results.

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 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 presented thus far have shown, this potential 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.¹⁵ 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 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 7 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

¹⁵See Appendix B for details.

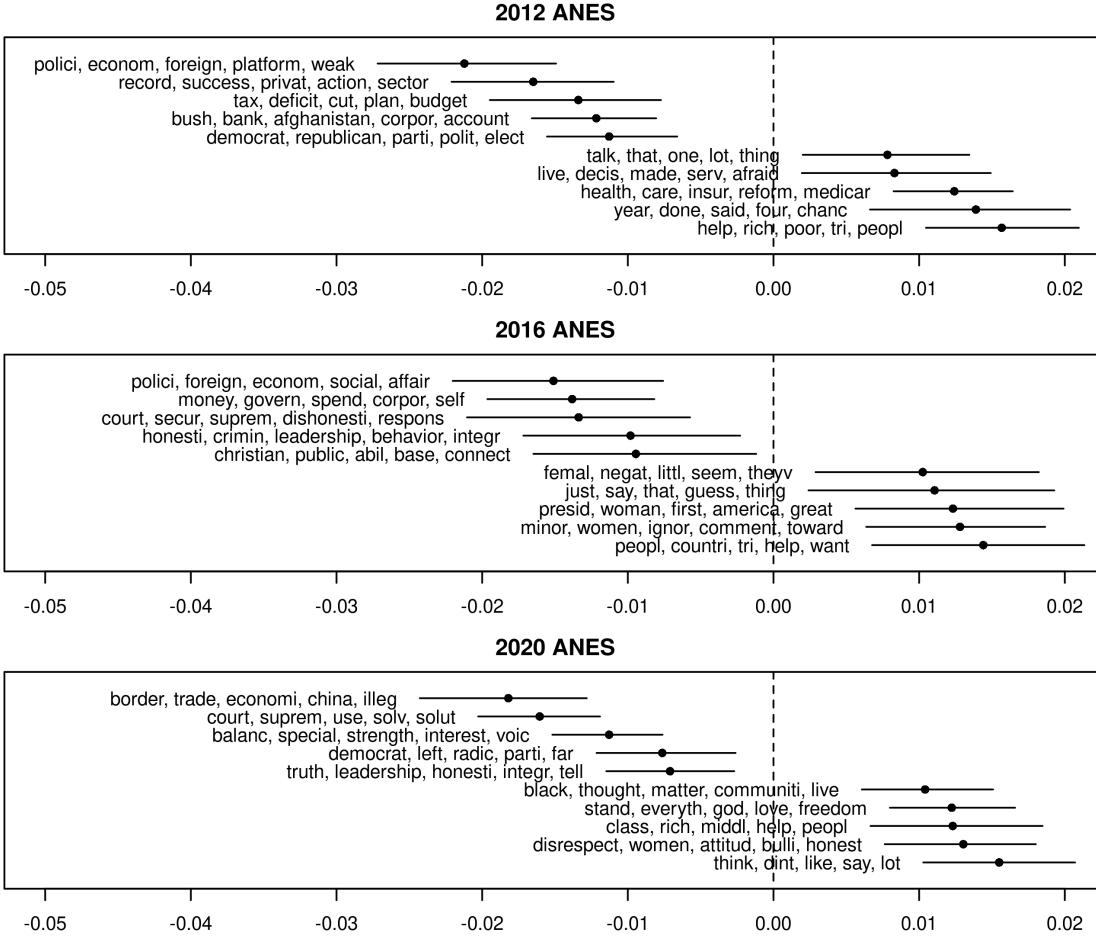


Figure 7: Gender differences in topic proportions in open-ended responses based on the structural topic model used to compute discursive sophistication (including 95% confidence intervals). Coefficients indicate the difference in predicted topic prevalence among women and men; positive values indicate higher prevalence among women. Labels are based on the five most frequent and exclusive (FREX) terms associated with each topic.

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, men are more likely to mention the topic revolving around terms like *tax*, *deficit*, and *cut*. Overall, across all three 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 specific considerations that are most salient to them.

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 political competence. 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 appropriately in their decision-making process. 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, our discipline lacks a 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 propor-

tion 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, a skeptic may still argue that while open-ended responses may provide useful insights, manual coding is still 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 (2018) 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.¹⁶

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.

¹⁶Package 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.
- 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. 2018. "Justifications and Citizen Competence in Direct Democracy: A Multilevel Analysis." *British Journal of Political Science* 48 (3): 787–806.
- 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. 1993. "Measuring political knowledge: Putting first things first." *American Journal of Political Science* 37 (4): 1179–1206.
- 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.
- Kraft, Patrick, and Kathleen Dolan. 2022. "Glass Half Full or Half Empty: Does Optimism about Women's Representation in Elected Office Matter?" *Journal of Women, Politics & Policy*: 1–13.
- 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.
- Lau, Richard R., and David P. Redlawsk. 2001. "Advantages and Disadvantages of Cognitive Heuristics in Political Decision Making." *American Journal of Political Science* 45 (4): 951-971.
- Lizotte, Mary-Kate, and Andrew H Sidman. 2009. "Explaining the gender gap in political knowledge." *Politics & Gender* 5 (02): 127–151.
- Lodge, Milton, and Charles S Taber. 2013. *The Rationalizing Voter*. Cambridge University Press.
- 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.
- Macdonald, Stuart Elaine, George Rabinowitz, and Ola Listhaug. 1995. "Political Sophistication and Models of Issue Voting." *British Journal of Political Science* 25 (4): 453-483.
- 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* 2 (2): 197–220.

- 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, 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.
- Vegetti, Federico, and Moreno Mancosu. 2020. "The Impact of Political Sophistication and Motivated Reasoning on Misinformation." *Political Communication* 37 (5): 678–695.
- 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. 2021. "Descriptive representation and political knowledge." *Politics, Groups, and Identities* 9 (1): 129–150.
- 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 Data Overview	1
I Open-Ended Items	1
II Conventional Knowledge Items	2
B Information on Discursive Sophistication Components	3
I Distribution of Word Counts & Proportion of Non-Response	3
II Preprocessing and Topic Model Specification	4
III Discursive Sophistication Components	5
C Robustness Checks	6
I PreText Analysis	6
II Discursive Sophistication for Varying Model Specifications	7
III Controlling for Personality and Verbal Skills	9
IV Personality and Verbal Skills as Predictors of Discursive Sophistication	10
V Discursive Sophistication and Uncertainty in Ideological Placements	11
VI Discursive Sophistication and Ideological Proximity Voting	13
D Tables of Model Estimates	14
I Figure 2: Effects of sophistication on turnout, political interest, internal efficacy, and external efficacy	14
II Figure 3: Expected information retrieval in the 2015 YouGov Study as a function of political sophistication	18
III Figure 6: The gender gap in political sophistication after controlling for common determinants.	19

Appendix A Data Overview

I Open-Ended Items

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

- On the issue of **gun legislation**, please outline the main arguments that come to mind *in favor and against* background checks for all gun sales, including at gun shows and over the Internet.
- On the issue of **abortion**, please outline the main arguments that come to mind *in favor and against* banning abortions after the 20th week of pregnancy.
- On the issue of **immigration**, please outline the main arguments that come to mind *in favor and against* providing a legal status for recipients of the Deferred Action for Childhood Arrivals (DACA) status.
- On the issue of **health 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.

2020, 2016, & 2012 American National Election Study (ANES): In addition to the original data collection as part of the CES, I apply the measurement approach using three separate waves 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. Here, discursive sophistication is evaluated using a set of 8 open-ended questions in which respondents were asked to list anything in particular that they like/dislike about the Democratic/Republican party as well as anything that might make them vote/not vote for either of the Presidential candidates.

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

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

II Conventional Knowledge Items

- *2018 CES*: Additive index containing 5 items (gun legislation, trade policy, DACA, health care, abortion).
- *2020 ANES*: Additive index containing 4 items (length of Senate term, federal government spending, majority in House, majority in Senate).
- *2016 ANES*: Additive index containing 4 items (length of Senate term, federal government spending, majority in House, majority in Senate).
- *2012 ANES*: Additive index containing 5 items (number of Presidential terms, size of budget deficit, length of Senate term, meaning of Medicare, federal government spending).
- *2015 YouGov*: Additive index containing 8 items (Speaker of the House, meaning of TPP, Chair of Federal Reserve Board, current unemployment rate, Presidential veto override, meaning of Common Core, leading source of electricity in US, majority in Senate).

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

I Distribution of Word Counts & Proportion of Non-Response

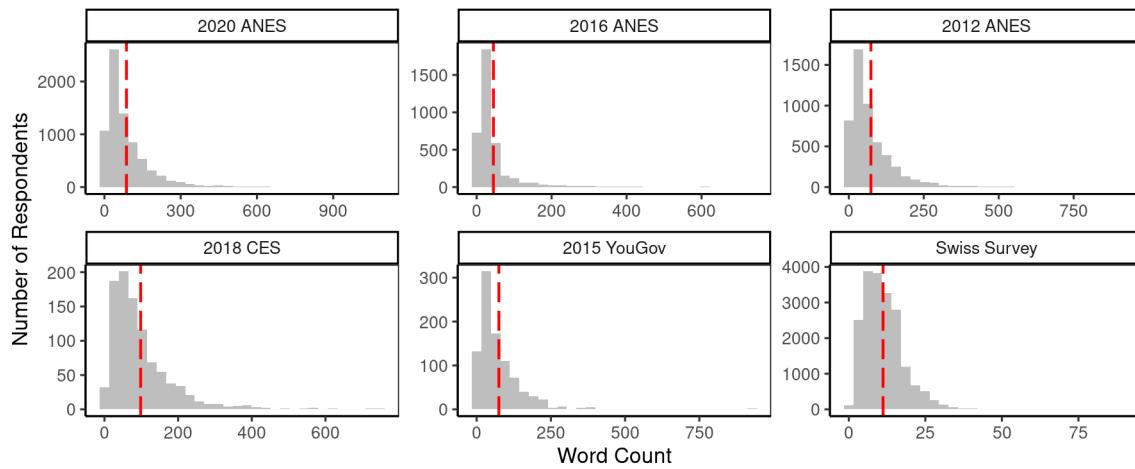


Figure B.1: Total word count across all open-ended responses for each survey participant. The dashed red lines indicate the average response lengths in each survey.

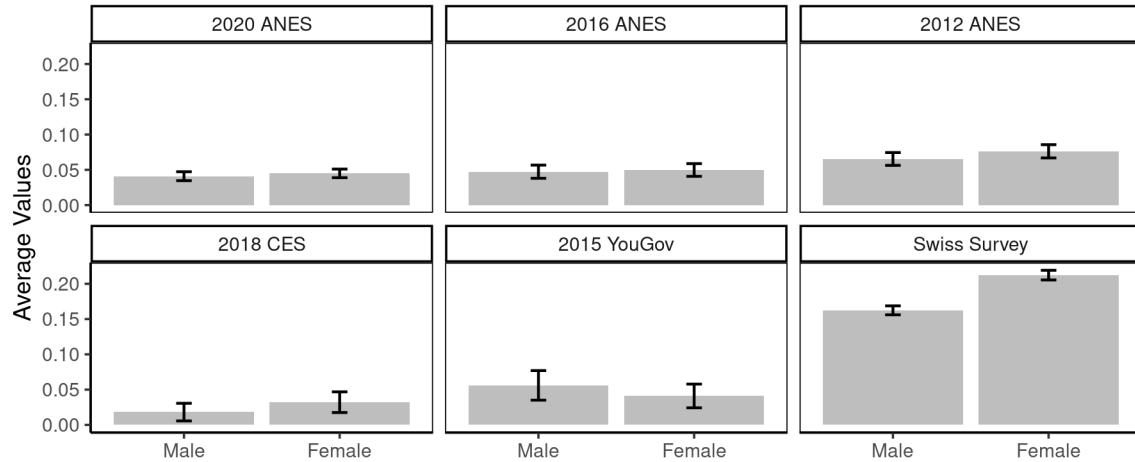


Figure B.2: Proportion of non-response comparing male and female survey participants (including 95% confidence intervals). Gender differences are only significant ($p < .05$) for the Swiss survey. Note, however, that respondents in the Swiss survey were only asked open-ended questions if they voted in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

II Preprocessing and Topic Model Specification

I rely on the structural topic model framework to extract and differentiate considerations mentioned by respondents. I follow the guidelines in [Roberts et al. \(2014\)](#) to preprocess our open-ended responses (lowercasing and stemming as well as removing stopwords, punctuation, numbers, and infrequent terms)¹⁷ and used age, gender, education, party identification, as well as an interaction between education and party identification as covariates for topic prevalence. With the exception of gender, this variable selection is equivalent to the procedure described in [Roberts et al. \(2014\)](#) for open-ended survey responses such as those included in the ANES. I set the number of topics to 25 in order to focus on the differentiation of broader considerations, but we replicate equivalent results with larger numbers of topics below. Figure B.3 displays the resulting topic proportions for all data sets included in the analyses along with the most frequent and exclusive (FREX) terms associated with each topic.

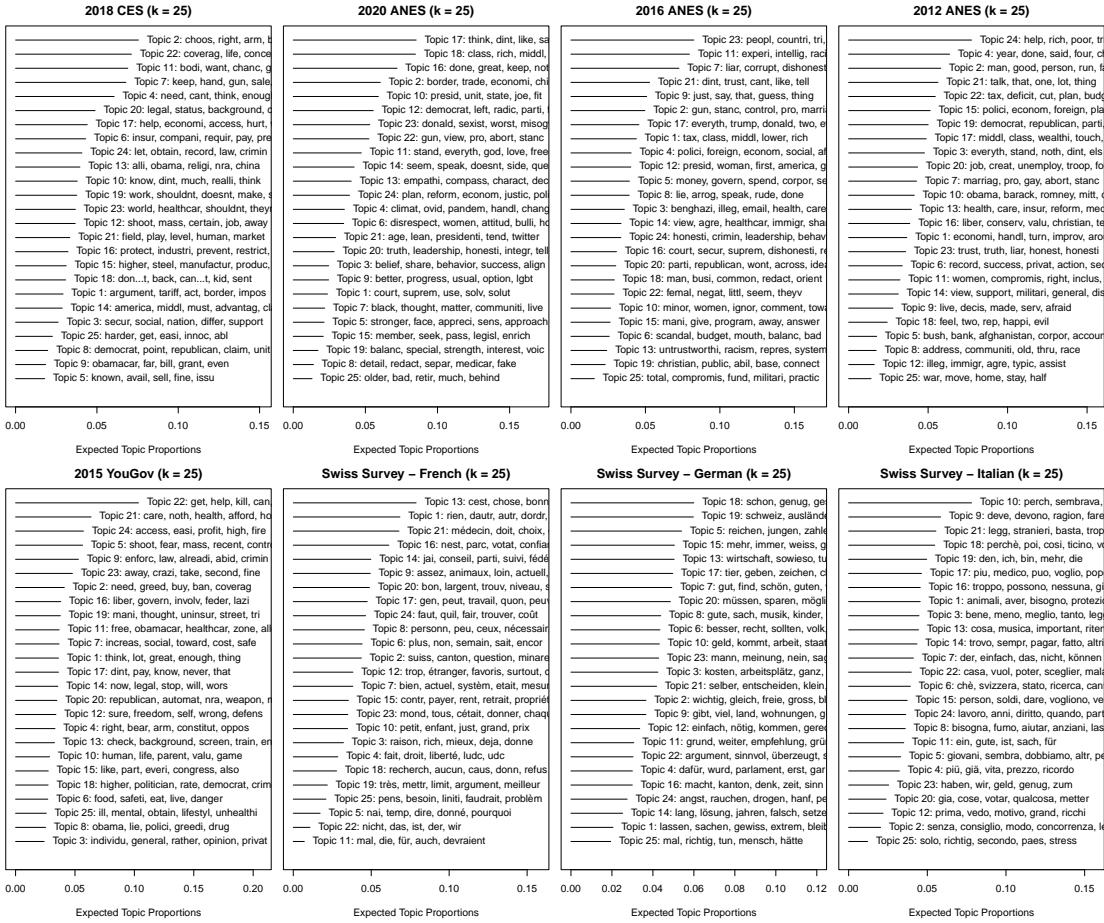


Figure B.3: Estimated topic proportions based on the structural topic model. See Appendix I for details on the model specification.

¹⁷Prior to applying these preprocessing steps, responses are cleaned by removing open-ended item non-response such as ‘don’t know’ and correcting spelling errors using an implementation of the Aspell spell-checking algorithm (www.aspell.net).

III Discursive Sophistication Components

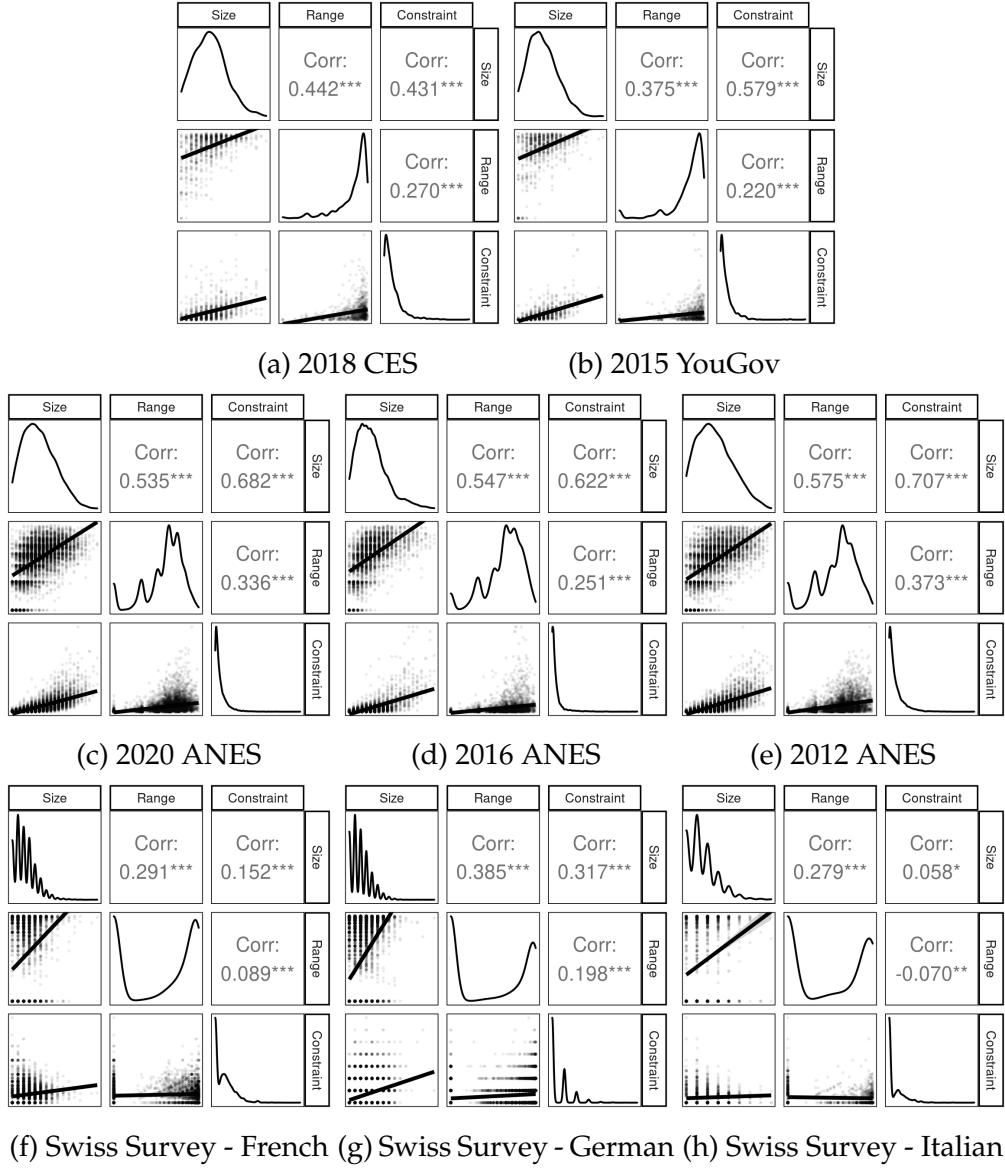


Figure B.4: Correlation matrix of individual components of discursive sophistication. The plots on the diagonal display univariate densities for each component. The panels in the lower triangular display the scatter plot of two measures as well as a linear fit.

Appendix C Robustness Checks

I PreText Analysis

The first component of discursive sophistication (*size*) relies on quantities extracted from structural topic models (Roberts et al., 2014). As with any other text-as-data approach, a necessary first step before estimating the topic model is to preprocess the raw text and convert it into a document term matrix (DTM, see for example Manning et al., 2008). Common preprocessing procedures include stemming and lowercasing, as well as the removal of numbers, punctuation, stopwords, and infrequent terms. However, topic models and other unsupervised learning techniques can be sensitive to these preprocessing choices. To address this issue, Denny and Spirling (2018) recommend that researchers compare DTMs under all possible preprocessing regimes. The authors propose *preText scores* as a measure to quantify the extent to which varying preprocessing regimes may yield unusual results compared to a baseline without any preprocessing. Following the procedure outlined in Denny and Spirling (2018), Figure C.1 displays the results of a linear model regressing preText scores resulting from all possible preprocessing regimes on each individual step for a subset of 500 open-ended responses in each of the surveys included in the analyses. Significant coefficients indicate that the topic model results may be sensitive to the respective preprocessing step.

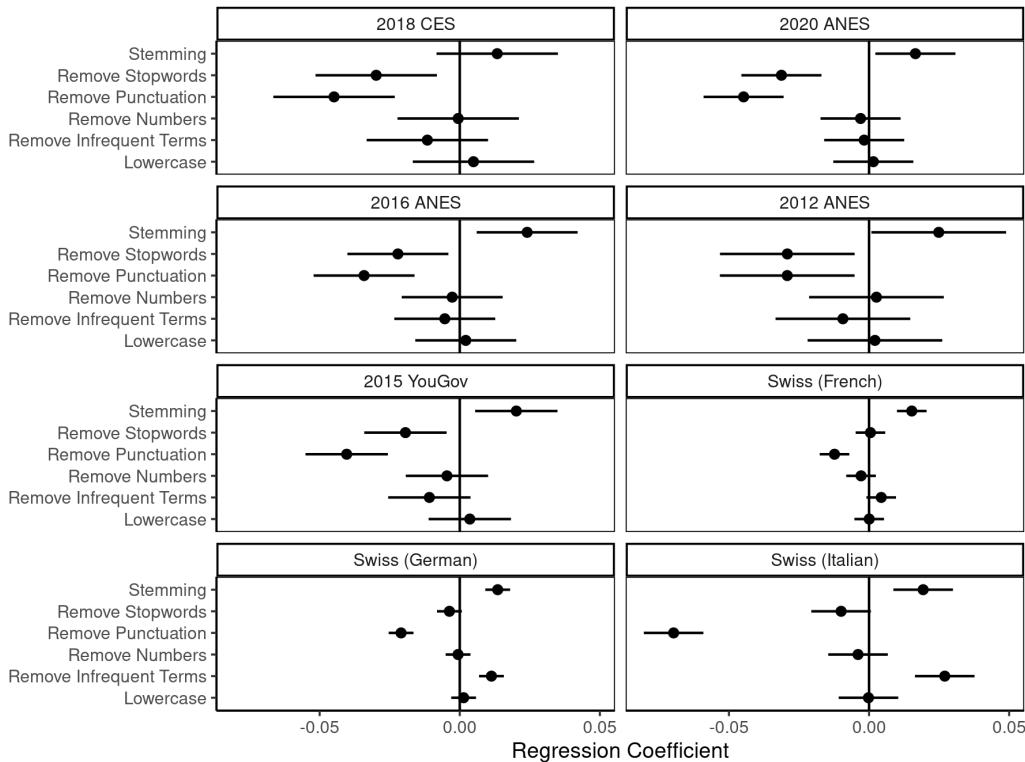


Figure C.1: PreText analysis of preprocessing decisions of open-ended responses across all datasets. Regression coefficients display the effects of each of the six preprocessing choices on the resulting preText score.

II Discursive Sophistication for Varying Model Specifications

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

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

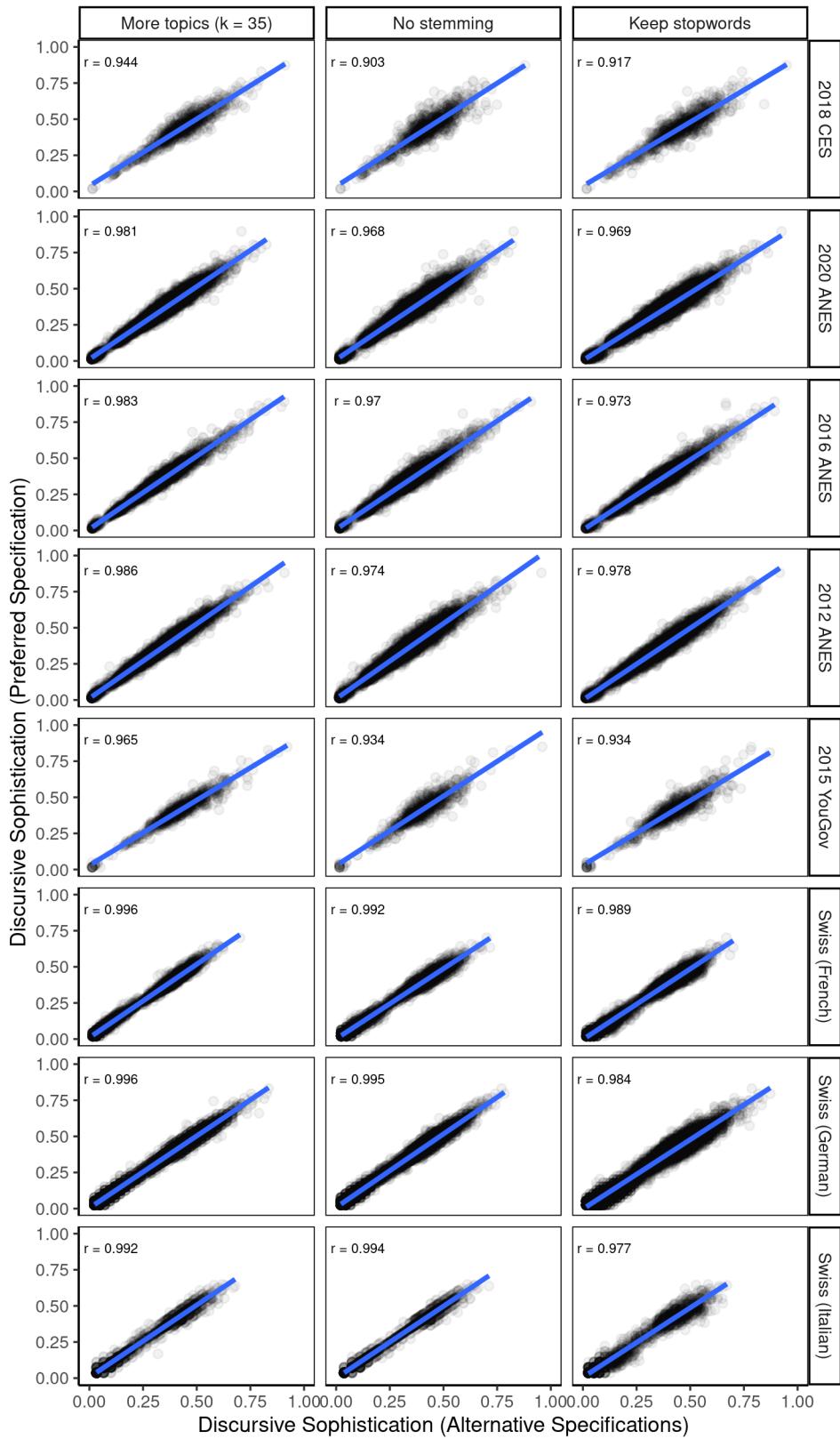


Figure C.2: Robustness of discursive sophistication measure for different preprocessing choices and topic model specifications.

III Controlling for Personality and Verbal Skills

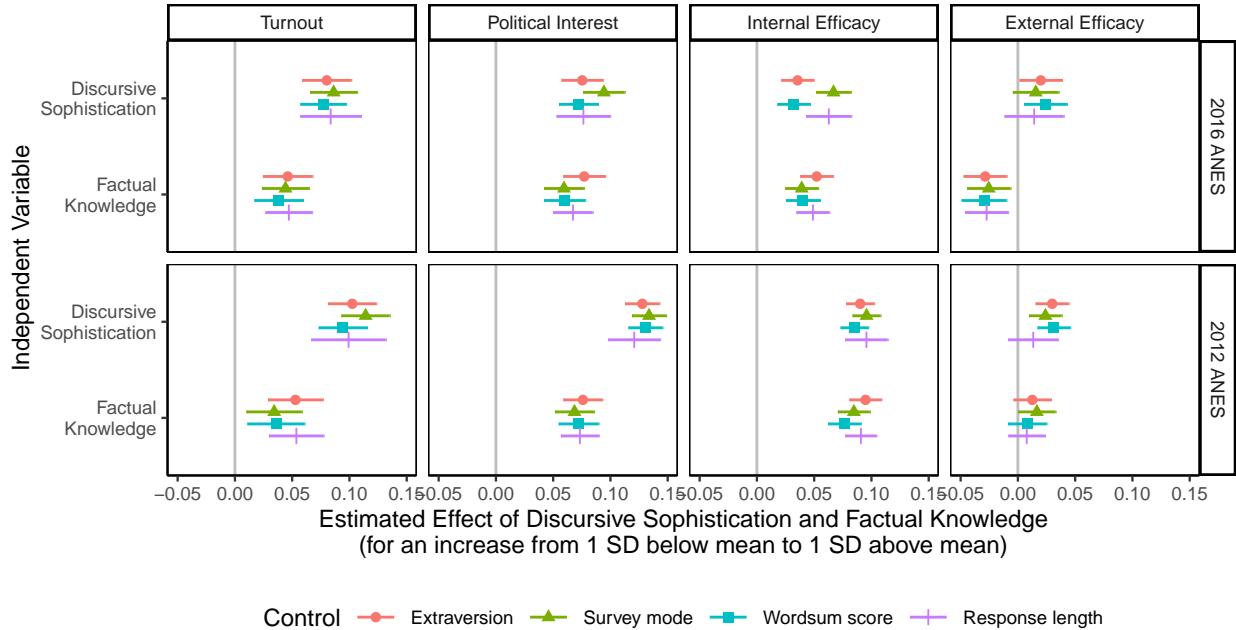


Figure C.3: Effects of sophistication on internal efficacy, external efficacy, non-conventional participation, and turnout in the 2012 and 2016 ANES. For each dependent variable, the figure displays the average marginal effects (AME) for each sophistication measure (including 95% confidence intervals).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

Description of additional control variables:

- *Personality*: Components of the Ten Item Personality Inventory (TIPI) measuring the “Big Five” personality traits measuring extraversion, openness to experience, being reserved.
- *Survey Mode*: Dichotomous indicator for face-to-face vs. online samples of the ANES surveys.
- *Wordsum vocabulary scores*: Modified version of the GSS wordsum vocabulary test consisting of 10 terms.
- *Response length*: Total number of words in the collection of open-ended responses by each individual.

IV Personality and Verbal Skills as Predictors of Discursive Sophistication

Table C.1: Personality, verbal skills, and survey mode as predictors of discursive sophistication and factual knowledge in the 2016 and 2012 ANES.

	<i>Dependent variable:</i>			
	Discursive Sophistication		Factual Knowledge	
	2016 ANES	2012 ANES	2016 ANES	2012 ANES
	(1)	(2)	(3)	(4)
Personality: Extraversion	0.049 (0.069)	-0.074 (0.061)	-0.051 (0.073)	-0.038 (0.053)
Personality: Openness to Experience	0.218** (0.078)	0.258*** (0.064)	0.085 (0.083)	-0.113* (0.055)
Personality: Reserved	-0.116* (0.058)	-0.128* (0.050)	-0.053 (0.061)	0.016 (0.044)
Verbal Skills (Wordsum score)	0.940*** (0.084)	1.116*** (0.070)	1.117*** (0.089)	1.165*** (0.061)
Survey Mode (Online)	-0.897*** (0.035)	-0.423*** (0.031)	0.232*** (0.037)	0.200*** (0.027)
Female	-0.060 (0.032)	-0.025 (0.027)	-0.127*** (0.034)	-0.240*** (0.023)
Age	0.0004 (0.001)	0.005*** (0.001)	0.006*** (0.001)	0.008*** (0.001)
Black	-0.071 (0.056)	0.110** (0.037)	0.137* (0.060)	-0.190*** (0.032)
Education: High School	0.122 (0.079)	0.120* (0.054)	0.087 (0.084)	0.215*** (0.047)
Education: Some College	0.342*** (0.076)	0.325*** (0.054)	0.097 (0.081)	0.336*** (0.046)
Education: Bachelor's Degree	0.554*** (0.081)	0.514*** (0.060)	0.157 (0.085)	0.507*** (0.052)
Education: Graduate Degree	0.586*** (0.085)	0.606*** (0.066)	0.218* (0.090)	0.582*** (0.057)
Household Income	0.207*** (0.060)	0.214*** (0.050)	0.206** (0.063)	0.361*** (0.043)
Church Attendance	0.020 (0.047)	0.080* (0.039)	0.018 (0.050)	0.015 (0.033)
Constant	-0.630*** (0.120)	-1.234*** (0.093)	-1.390*** (0.127)	-1.548*** (0.081)
Observations	3,018	4,696	3,014	4,696
R ²	0.266	0.179	0.138	0.343

Note:

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

V Discursive Sophistication and Uncertainty in Ideological Placements

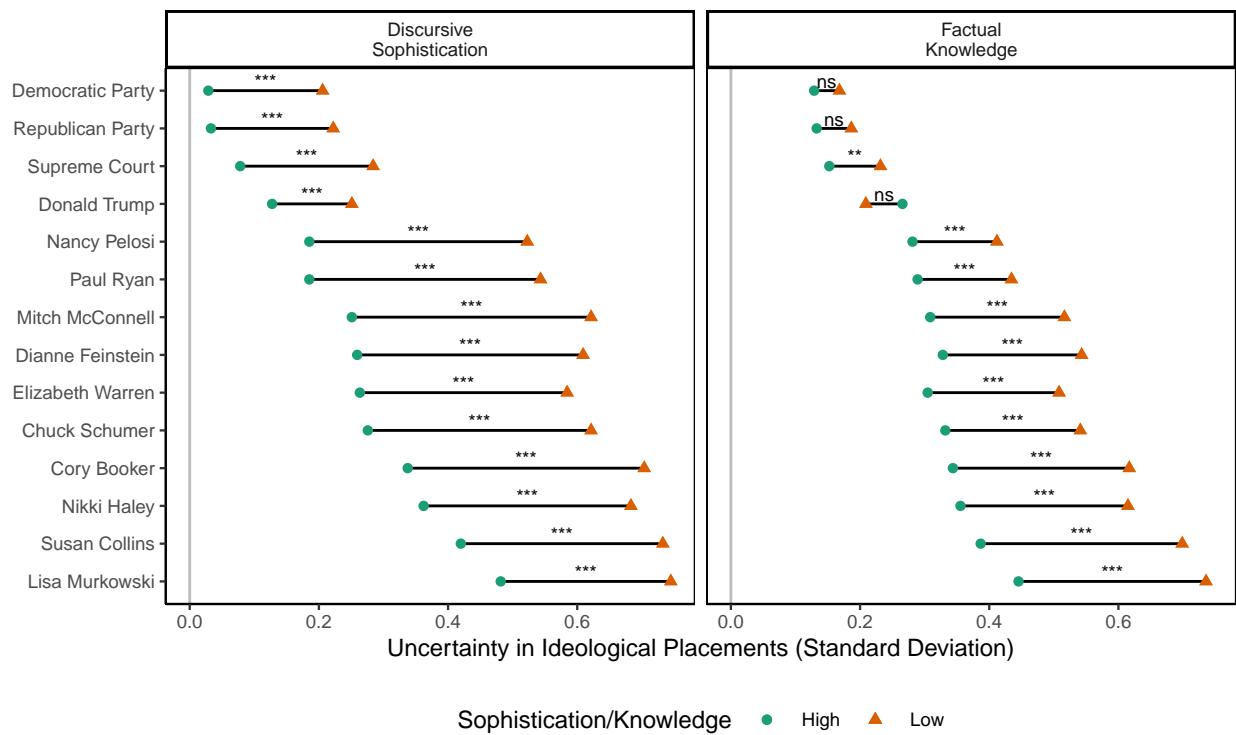


Figure C.4: Insert caption here.

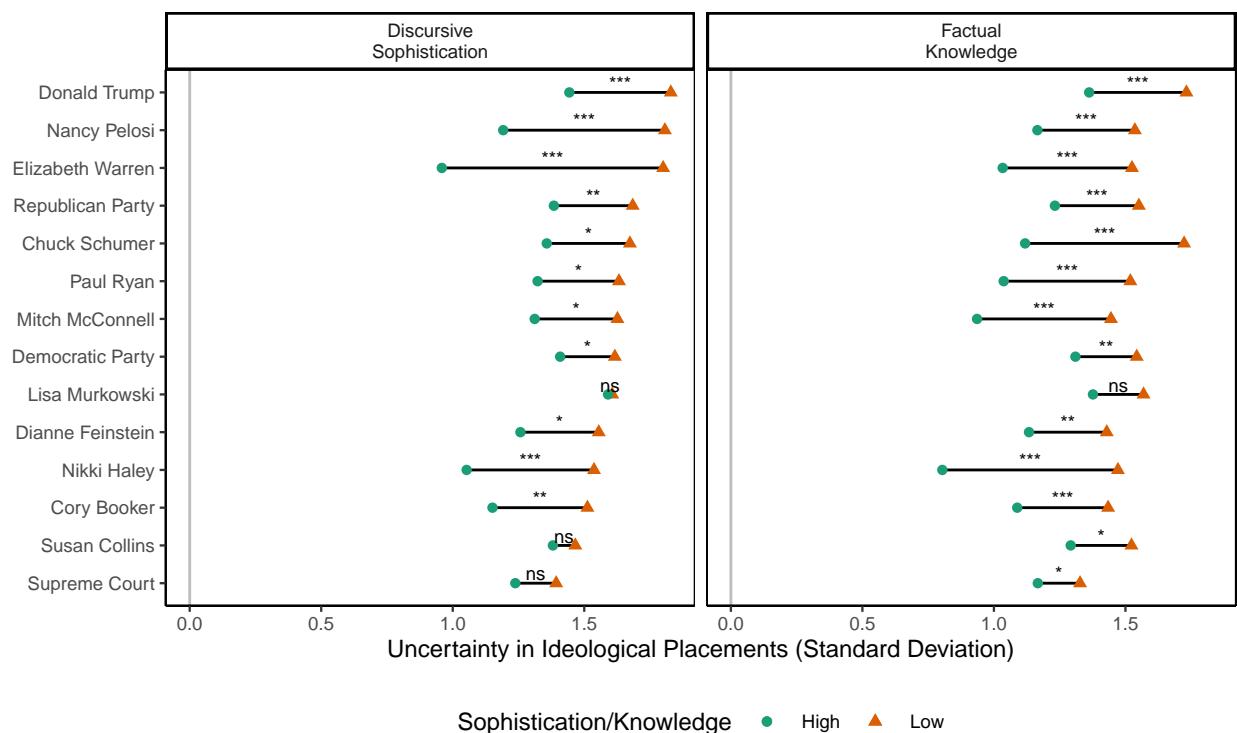


Figure C.5: Insert caption here.

VI Discursive Sophistication and Ideological Proximity Voting

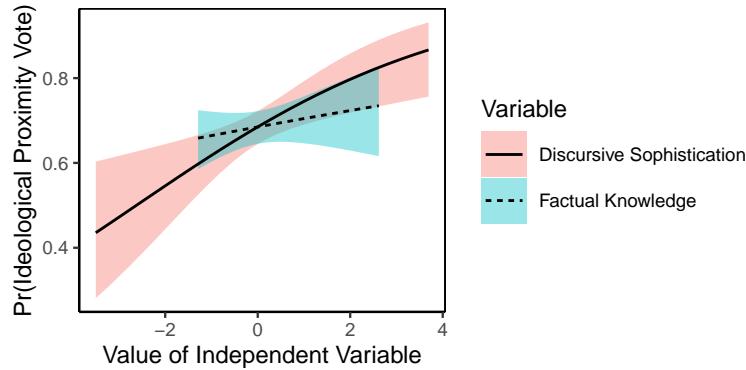


Figure C.6: Insert caption here.

Table C.2: Logistic regression predicting ideological proximity-based voting for US Senators in the 2018 CES. Standard errors in parentheses. Estimates are used for Figure C.6.

	<i>Dependent variable:</i> Ideological Proximity Vote
Discursive Soph.	0.296** (0.097)
Factual Knowledge	0.093 (0.101)
Female	−0.226 (0.188)
Age	0.016** (0.005)
Black	−0.177 (0.328)
College Degree	0.408* (0.200)
Household Income	0.322 (0.441)
Church Attendance	0.104 (0.274)
Constant	−0.209 (0.344)
Observations	611
Akaike Inf. Crit.	746.658

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

Appendix D Tables of Model Estimates

I Figure 2: 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 in model (2), (5), (8), and (11) are used for Figure 2 in the main text.

	Dependent variable:											
	Turnout			Political Interest			Internal Efficacy			External Efficacy		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Discursive Soph.	0.411*** (0.081)	0.282** (0.091)	0.272** (0.093)	0.049*** (0.009)	0.033*** (0.009)	0.033*** (0.009)	0.050*** (0.008)	0.039*** (0.008)	0.039*** (0.008)	-0.001 (0.009)	-0.009 (0.009)	-0.009 (0.009)
Factual Knowledge	0.439*** (0.090)	0.286** (0.103)	0.286** (0.103)	0.083*** (0.009)	0.066*** (0.009)	0.067*** (0.009)	0.073*** (0.008)	0.062*** (0.008)	0.062*** (0.008)	0.017 (0.009)	0.018 (0.010)	0.017 (0.010)
Disc. X Factual				-0.046 (0.094)			-0.006 (0.008)			-0.002 (0.007)		0.008 (0.009)
Female		-0.518** (0.190)	-0.519** (0.190)		-0.081*** (0.017)	-0.081*** (0.017)		-0.050** (0.016)	-0.050** (0.016)	0.020 (0.018)	0.020 (0.018)	
Age			0.044*** (0.006)	0.044*** (0.006)		0.004*** (0.0005)	0.004*** (0.0005)		0.002*** (0.0004)	0.002*** (0.0004)	-0.0004 (0.001)	-0.0004 (0.001)
Black		-0.595* (0.266)	-0.600* (0.267)		-0.041 (0.028)	-0.041 (0.028)		0.030 (0.026)	0.030 (0.026)	-0.026 (0.030)	-0.026 (0.030)	
College Degree			0.641** (0.206)	0.649** (0.207)		0.031 (0.018)	0.032 (0.018)		0.039* (0.017)	0.039* (0.017)	0.037 (0.020)	0.036 (0.020)
Household Income			1.087* (0.461)	1.086* (0.461)		0.158*** (0.041)	0.158*** (0.041)		0.114** (0.038)	0.114** (0.038)	0.027 (0.044)	0.028 (0.044)
Church Attendance			0.657* (0.282)	0.654* (0.282)		0.060* (0.025)	0.059* (0.025)		0.007 (0.023)	0.007 (0.023)	0.086** (0.026)	0.086** (0.026)
Constant	1.207*** (0.082)	-1.188*** (0.340)	-1.175*** (0.341)	0.623*** (0.008)	0.398*** (0.032)	0.401*** (0.032)	0.607*** (0.007)	0.456*** (0.029)	0.457*** (0.029)	0.342*** (0.008)	0.299*** (0.034)	0.295*** (0.034)
Observations	943	849	849	941	848	848	941	847	847	941	848	848
R ²				0.152	0.278	0.279	0.163	0.225	0.225	0.004	0.025	0.026
Akaike Inf. Crit.	991.648	793.140	794.898									

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 in model (2), (5), (8), and (11) are used for Figure 2 in the main text.

	Dependent variable:											
	Turnout			Political Interest			Internal Efficacy			External Efficacy		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Discursive Soph.	0.527*** (0.039)	0.388*** (0.042)	0.384*** (0.045)	0.053*** (0.003)	0.052*** (0.003)	0.052*** (0.003)	0.009*** (0.002)	0.010*** (0.002)	0.010*** (0.002)	0.016*** (0.003)	0.005 (0.003)	0.006 (0.003)
Factual Knowledge	0.396*** (0.041)	0.269*** (0.047)	0.265*** (0.049)	0.059*** (0.003)	0.041*** (0.003)	0.041*** (0.003)	0.015*** (0.002)	0.012*** (0.002)	0.012*** (0.002)	0.024*** (0.003)	0.020*** (0.004)	0.019*** (0.004)
Disc. X Factual				-0.011 (0.044)			-0.002 (0.003)			-0.001 (0.002)		0.010** (0.003)
Female	0.442*** (0.085)	0.442*** (0.085)			-0.042*** (0.006)	-0.042*** (0.006)			-0.019*** (0.004)	-0.019*** (0.004)	0.028*** (0.007)	0.028*** (0.007)
Age	0.030*** (0.003)	0.030*** (0.003)			0.003*** (0.0002)	0.003*** (0.0002)			0.0003* (0.0001)	0.0003* (0.0001)	-0.00003 (0.0002)	-0.00002 (0.0002)
Black	0.251 (0.137)	0.251 (0.137)			0.021* (0.010)	0.021* (0.010)			-0.010 (0.008)	-0.010 (0.008)	0.029* (0.012)	0.029* (0.012)
College Degree	0.711*** (0.100)	0.713*** (0.100)			0.015* (0.006)	0.015* (0.006)			-0.014** (0.005)	-0.014** (0.005)	0.066*** (0.007)	0.066*** (0.007)
Household Income	1.300*** (0.142)	1.301*** (0.142)			0.017 (0.009)	0.017 (0.009)			0.011 (0.007)	0.011 (0.007)	0.056*** (0.011)	0.057*** (0.011)
Church Attendance	0.559*** (0.143)	0.558*** (0.144)			-0.012 (0.009)	-0.012 (0.009)			-0.031*** (0.007)	-0.031*** (0.007)	0.014 (0.010)	0.015 (0.010)
Constant	2.165*** (0.043)	-0.634*** (0.154)	-0.634*** (0.154)	0.680*** (0.003)	0.508*** (0.011)	0.509*** (0.011)	0.605*** (0.011)	0.608*** (0.002)	0.608*** (0.009)	0.355*** (0.009)	0.273*** (0.003)	0.269*** (0.013)
Observations	6,694	6,370	6,370	7,323	6,964	6,964	6,633	6,317	6,317	6,632	6,315	6,315
R ²				0.126	0.185	0.185	0.013	0.022	0.022	0.015	0.042	0.044
Akaike Inf. Crit.	4,547.295	3,966.170	3,968.102									

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 in model (2), (5), (8), and (11) are used for Figure 2 in the main text.

	Dependent variable:											
	Turnout			Political Interest			Internal Efficacy			External Efficacy		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Discursive Soph.	0.546*** (0.055)	0.448*** (0.058)	0.446*** (0.061)	0.049*** (0.004)	0.038*** (0.004)	0.039*** (0.004)	0.030*** (0.004)	0.019*** (0.004)	0.019*** (0.004)	0.021*** (0.005)	0.011* (0.005)	0.012* (0.005)
Factual Knowledge	0.339*** (0.053)	0.258*** (0.057)	0.255*** (0.061)	0.049*** (0.004)	0.034*** (0.004)	0.034*** (0.004)	0.036*** (0.004)	0.026*** (0.004)	0.026*** (0.004)	-0.006 (0.004)	-0.014** (0.005)	-0.014** (0.005)
Disc. X Factual		-0.006 (0.057)			0.003 (0.004)			0.002 (0.004)			0.006 (0.005)	
Female	0.181 (0.112)	0.180 (0.112)		-0.064*** (0.008)	-0.063*** (0.008)		-0.055*** (0.007)	-0.054*** (0.007)		-0.001 (0.009)	-0.001 (0.009)	
Age	0.022*** (0.003)	0.022*** (0.003)		0.004*** (0.0002)	0.004*** (0.0002)		0.001** (0.0002)	0.001** (0.0002)		0.00001 (0.0003)	0.00000 (0.0003)	
Black	1.044*** (0.233)	1.045*** (0.233)		0.022 (0.015)	0.022 (0.015)		0.042** (0.012)	0.042** (0.012)		-0.015 (0.016)	-0.015 (0.016)	
College Degree	0.674*** (0.137)	0.675*** (0.137)		0.061*** (0.009)	0.060*** (0.009)		0.070*** (0.008)	0.070*** (0.008)		0.059*** (0.010)	0.058*** (0.010)	
Household Income	0.691*** (0.202)	0.690*** (0.202)		0.033* (0.015)	0.034* (0.015)		0.042** (0.013)	0.043*** (0.013)		0.060*** (0.017)	0.061*** (0.017)	
Church Attendance	1.057*** (0.188)	1.057*** (0.188)		0.008 (0.012)	0.008 (0.012)		-0.010 (0.010)	-0.010 (0.010)		0.080*** (0.013)	0.079*** (0.013)	
Constant	2.183*** (0.058)	0.085 (0.197)	0.085 (0.197)	0.637*** (0.004)	0.440*** (0.016)	0.440*** (0.016)	0.541*** (0.004)	0.485*** (0.014)	0.485*** (0.014)	0.397*** (0.014)	0.316*** (0.005)	0.315*** (0.018)
Observations	3,671	3,562	3,562	3,694	3,582	3,582	3,190	3,104	3,104	3,192	3,106	3,106
R ²				0.073	0.157	0.158	0.056	0.113	0.113	0.007	0.040	0.041
Akaike Inf. Crit.	2,470.276	2,256.640	2,258.629									

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 in model (2), (5), (8), and (11) are used for Figure 2 in the main text.

	Dependent variable:											
	Turnout			Political Interest			Internal Efficacy			External Efficacy		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Discursive Soph.	0.501*** (0.041)	0.408*** (0.044)	0.418*** (0.044)	0.069*** (0.004)	0.065*** (0.004)	0.065*** (0.004)	0.049*** (0.003)	0.046*** (0.003)	0.045*** (0.003)	0.022*** (0.003)	0.015*** (0.004)	0.015*** (0.004)
Factual Knowledge	0.375*** (0.041)	0.212*** (0.048)	0.233*** (0.050)	0.055*** (0.004)	0.036*** (0.004)	0.037*** (0.004)	0.052*** (0.003)	0.046*** (0.003)	0.046*** (0.003)	-0.002 (0.004)	0.003 (0.004)	0.004 (0.004)
Disc. X Factual			0.060 (0.045)			0.009* (0.004)			0.006 (0.003)			0.009** (0.004)
Female	0.071 (0.084)	0.074 (0.084)		-0.064*** (0.007)	-0.064*** (0.007)		-0.052*** (0.006)	-0.052*** (0.006)		0.017* (0.007)	0.017* (0.007)	
Age	0.027*** (0.003)	0.027*** (0.003)		0.003*** (0.002)	0.003*** (0.002)		0.00001 (0.0002)	0.00002 (0.0002)		-0.0003 (0.0002)	-0.0003 (0.0002)	
Black	0.804*** (0.120)	0.801*** (0.120)		0.041*** (0.010)	0.041*** (0.010)		0.054*** (0.008)	0.054*** (0.008)		0.084*** (0.009)	0.084*** (0.009)	
College Degree	0.475*** (0.109)	0.463*** (0.110)		0.031*** (0.008)	0.029*** (0.008)		0.032*** (0.007)	0.031*** (0.007)		0.036*** (0.008)	0.034*** (0.008)	
Household Income	1.101*** (0.153)	1.105*** (0.153)		0.001 (0.013)	0.001 (0.013)		0.021* (0.011)	0.021* (0.011)		0.011 (0.013)	0.012 (0.013)	
Church Attendance	0.655*** (0.127)	0.657*** (0.127)		0.010 (0.010)	0.011 (0.010)		0.005 (0.008)	0.006 (0.008)		0.049*** (0.010)	0.050*** (0.010)	
Constant	1.662*** (0.042)	-0.597*** (0.148)	-0.610*** (0.148)	0.607*** (0.004)	0.477*** (0.014)	0.474*** (0.014)	0.569*** (0.003)	0.564*** (0.011)	0.562*** (0.011)	0.391*** (0.011)	0.349*** (0.003)	0.345*** (0.013)
Observations	4,847	4,714	4,714	5,164	5,002	5,002	5,154	4,994	4,994	5,140	4,983	4,983
R ²				0.131	0.176	0.177	0.128	0.155	0.156	0.008	0.038	0.039
Akaike Inf. Crit.	4,192.410	3,799.655	3,799.819									

Note:

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

II Figure 3: Expected information retrieval in the 2015 YouGov Study as a function of political sophistication

Table D.5: Linear regressions predicting information retrieval in the 2015 YouGov study. Standard errors in parentheses. Estimates are used for Figure 3 in the main text.

	<i>Dependent variable:</i>		
	Information Retrieval		
	(1)	(2)	(3)
Discursive Soph.	0.039*** (0.006)	0.033*** (0.006)	0.032*** (0.006)
Factual Knowledge	0.030*** (0.006)	0.035*** (0.007)	0.034*** (0.007)
Disc. X Factual			-0.008 (0.006)
Female		0.043*** (0.012)	0.043*** (0.012)
Age		0.001* (0.0004)	0.001* (0.0004)
Black		-0.027 (0.021)	-0.026 (0.021)
College Degree		0.017 (0.014)	0.017 (0.014)
Household Income		-0.061* (0.030)	-0.060* (0.030)
Church Attendance		-0.059*** (0.017)	-0.059*** (0.017)
Constant	0.828*** (0.005)	0.807*** (0.021)	0.811*** (0.021)
Observations	918	792	792
R ²	0.100	0.137	0.139

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

III Figure 6: The gender gap in political sophistication after controlling for common determinants.

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

	Dependent variable:							
	2018 CES (1)	2020 ANES (2)	2016 ANES (3)	Discursive Sophistication				Italian (8)
				2012 ANES (4)	2015 YouGov (5)	French (6)	German (7)	
Female	-0.015 (0.070)	-0.051* (0.023)	-0.029 (0.033)	-0.0001 (0.027)	0.075 (0.072)	0.024 (0.029)	-0.051** (0.018)	0.039 (0.048)
Age	0.006** (0.002)	0.004*** (0.001)	0.003** (0.001)	0.005*** (0.001)	0.007** (0.002)	0.002* (0.001)	-0.001* (0.001)	0.001 (0.002)
Black	-0.321** (0.115)	-0.283*** (0.041)	-0.148** (0.057)	0.021 (0.036)	-0.448*** (0.130)			
College Degree	0.404*** (0.073)	0.436*** (0.024)	0.407*** (0.035)	0.460*** (0.031)	0.165 (0.085)	-0.093** (0.029)	0.005 (0.018)	-0.017 (0.048)
Household Income	0.124 (0.171)	0.446*** (0.038)	0.366*** (0.059)	0.409*** (0.049)	-0.007 (0.182)			
Church Attendance	0.052 (0.101)	0.040 (0.035)	0.074 (0.048)	0.074 (0.039)	-0.060 (0.102)			
Constant	-0.464*** (0.130)	-0.620*** (0.045)	-0.503*** (0.062)	-0.612*** (0.051)	-0.347** (0.130)	-0.060 (0.058)	0.096** (0.035)	-0.063 (0.097)
Observations	849	6,965	3,582	5,004	792	4,994	12,465	1,840
R ²	0.068	0.109	0.076	0.092	0.036	0.004	0.001	0.001

Note:

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

Table D.7: Linear regressions predicting factual knowledge in the CES, ANES, and YouGov study. Estimates are used for Figure 6 in the main text.

	Dependent variable:				
	2018 CES (1)	2020 ANES (2)	2016 ANES (3)	Factual Knowledge	
				2012 ANES (4)	2015 YouGov (5)
Female	-0.150* (0.067)	-0.352*** (0.022)	-0.095** (0.032)	-0.253*** (0.024)	-0.589*** (0.060)
Age	0.003 (0.002)	0.012*** (0.001)	0.008*** (0.001)	0.012*** (0.001)	0.012*** (0.002)
Black	-0.284* (0.111)	-0.172*** (0.039)	-0.036 (0.056)	-0.373*** (0.032)	-0.318** (0.108)
College Degree	0.508*** (0.071)	0.437*** (0.023)	0.211*** (0.034)	0.429*** (0.027)	0.414*** (0.071)
Household Income	0.254 (0.165)	0.394*** (0.037)	0.441*** (0.058)	0.653*** (0.043)	0.869*** (0.152)
Church Attendance	-0.320** (0.098)	-0.020 (0.034)	-0.085 (0.047)	-0.044 (0.034)	-0.190* (0.085)
Constant	-0.193 (0.125)	-0.803*** (0.043)	-0.605*** (0.061)	-0.747*** (0.045)	-0.491*** (0.109)
Observations	849	6,965	3,582	5,004	792
R ²	0.112	0.172	0.067	0.239	0.275

Note:

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

Response to Reviewers

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

The following memo details revisions I have made to my manuscript in light of the comments and suggestions I received from three reviewers. I thank each of the reviewers for their support of the project and the comments/suggestions provided. I have addressed each of their comments directly in the revised manuscript and/or by providing additional analyses in the appendix. For instance, I have done XXX. I believe the comments and suggestions have helped us produce a stronger revised manuscript and hope that the reviewers and editor agree.

Reviewer 1

This is a promising paper. In it, the author makes a convincing case that a key aspect of citizen sophistication, discursive sophistication, can be measured via automated text analysis of responses to open-ended survey items, and that the resulting measures 1) are only modestly correlated with conventional civics knowledge scales, 2) generally exert stronger influence on criterion variables than does civics knowledge, and 3) show no sign of the gender gap we typically observe on knowledge measures. This is a lot, especially given that analyses make use of data from multiple surveys. The paper is very interesting and largely persuasive. I see it as having good prospects for eventual publication.

My two concerns are somewhat broad, and each relates to aspects of how the paper is framed and what it ultimately accomplishes. Neither of these should be terribly difficult for the author to address.

First, the juxtaposition between the new measure and civics knowledge measures is overdone. At several points, the paper is framed as if to suggest that there is an implied competition between the new measure and knowledge scales. But there is no such competition either conceptually or empirically. Conceptually, there is no reason that political sophistication cannot include both an analytical, integrative component (as represented by the current paper's text-based measures) and an information raw material component (as represented by civics knowledge). Empirically, we see in the current paper's Figure 2 that the two are only modestly correlated, and in Figure 3 that both produce effects on typical political dependent variables (although it does not appear that both variables are included as IVs at the same time; they should be, at least as one specification in the appendix models). This all suggests that the two variables are complements, not competitors. To me, this is a better way to frame the paper: political scientists have obsessed about knowledge, but there is more to sophistication than knowledge; we can add a good measure of a different aspect of sophistication through automated text coding, and doing so deepens our understanding of the importance of sophistication, and does so in a manner that suggests an absence of a gender gap.

It also is possible that the text measure and the civics measure operate in conjunction with one another. I recommend exploring whether they interact. If they do, two different patterns seem sensible. First, there could be weak or null main effects, but a significant positive interaction, in models with the political DVs. This would suggest that knowledge and discursive sophistication are individually insufficient to produce positive effects, but

that those effects are achieved in combination. In other words, we get the most civic engagement when a person has a good baseline factual understanding of politics along with the cognitive sophistication to connect the dots and form cohesive, well-thought opinions. Second, and very different, there could be positive main effects, but a significant negative interaction. This would suggest that either knowledge or discursive sophistication is sufficient in itself to produce good behavioral effects, but that there is a diminishing return when they operate in combination. These tests, in turn, could be broken out by gender. That could be very valuable to explore, because it potentially could show that discursive sophistication not only does not exhibit a gender gap, but that it acts to reduce the effects of the standard gender gap in knowledge.

Second, there is room to do a more thorough job of differentiating between discursive sophistication as a general phenomenon and discursive political sophistication. This is mentioned briefly in the conclusion, but more discussion would be good. This issue harkens back to Nie et al.'s *The Changing American Voter* and Eric Smith's *The Unchanging American Voter*, with Smith showing that the supposed change observed by Nie et al. mostly involved verbosity, or general communication skill, rather than anything that was specifically political. The current paper has an appendix table in which discursive sophistication is regressed on some demographic variables. This is a start, but I would recommend 1) when discursive sophistication is the DV, adding a fuller representation of education as an IV, and including whatever measure of verbal skill is available, along with anything else, such as cognitive ability, that the surveys might include. In fact, civics knowledge also could be included. That way, the models would show how much of discursive sophistication is and is not accounted for by conventional predictors (the multivariate models with discursive sophistication as an IV also should include as covariates anything that is associated with both discursive sophistication and the DV), and 2) addressing this at the conceptual level in the paper. Does the discursive measure merely show that some people are better than others at formulating and communicating arguments in general, or is there something distinctly political about it? This, of course, links to my first point about sophistication vs. knowledge: it potentially is the case that the author's measure captures an important general aspect of sophistication, whereas civics knowledge contributes a political dimension.

Some minor points:

1. I would change the first part of the title. Discursive sophistication is not about "knowing" in the conventional sense, it is about sophistication in ability. Changing the title would help with decoupling discursive sophistication from knowledge, thereby improving the paper's framing.
2. Luskin's 1990 Political Behavior paper is worth a look.
3. It is Delli Carpini, not Carpini. "Delli" is part of the last name, not a middle name.
4. Figure 1 is not necessary.
5. I'm pretty sure Mondak (2001) was the first to make the point about coding problems on open-ended knowledge items (that's what is noted in Gibson and Caldeira 2009).

Reviewer 2

This manuscript proposes a novel way to measure the complex and polysemic concept of political sophistication. Rather than using typical survey items measuring factual knowl-

edge focusing on the accretion of electoral and partisan facts (and with closed ended format implying the selection of the correct answer among a list of potential answers), the paper proposes a measure of what it calls: "discursive sophistication" based on how people discuss their political preferences in open ended survey responses.

The piece is well written, and the findings are interesting. Moreover, I recognize that the authors have done an excellent job of situating their project in the literature and of making a case for their contribution. I am also impressed by the amount of empirical work presented in such an elegant way. However, in my view at present the contribution is methodological. In what follows I provide a number of comments and suggestions to improve the piece:

1-Regarding the discussion about the limits and problems associated to the measurement of political sophistication with factual political knowledge survey items, I think it is well focused and refined but there are two relevant factors that the paper does not discuss:

a-The inclination of conventional knowledge survey items to focus on the measurement of the capacity of citizens to remember facts. This implies giving priority to their declarative memory, while leaving aside their procedural memory. At present this piece does not dialog with this relevant literature.

b-The relevance of the format (that is true/false; Multiple choice, Open ended, etc.) to obtain different estimations of peoples' levels of political sophistication. There is a debate in the literature about this question too. The paper could benefit from the main results of this debate comparing different formats across conventional factual knowledge items. There is a recent paper at SSQ that deals with this (<https://onlinelibrary.wiley.com/doi/full/10.1111/ssqu.12822>)

2-Regarding the discussion about the potential sources of measurement bias influencing the apparent gender gap, I have also missed the relevance of format (that is true/false; Multiple choice, Open ended, etc.) and the temporal dimension of the questions. If I remember well two articles in APSR 2014 (Barabas et al) and P&G 2018 (Ferrin et al) discuss this point.

3-The whole idea about discursive sophistication is solidly based on a classic contribution to the field (Converse 1964 and Luskin 1987), or what is known in the literature as the structure of belief systems. However, at present I think that the contribution of the manuscript is mainly empirical. Or is there something new with respect to these two studies that the present manuscript advance? If such is the case, then the author(s) should make it explicit to convince the reader.

4-In the discussion about the definition of discursive sophistication (distinguishing three dimensions: size, range and constraint) it is not clear how each of these dimensions are linked to the concept that the author(s) intend(s) to measure: the extent to which citizens are able to understand the functioning of institutions, the performance of the incumbent government, and the actions of the main political actors. This is the only way for people to assess their interest as individuals and as members of groups.

5-Regarding the empirics, the paper uses existing survey evidence containing batteries of open-ended questions about very different specific topics such as gun legislation, health care, immigration, preferences for party and candidates. I wonder if these are the most adequate type of topics to measure discursive sophistication, given their partisan

roots. In other words: how ideology or partisanship might be affecting the main findings reported in this paper?

The piece does discuss the univariate densities across the three indicators: discursive sophistication has always a close to normal distribution while factual knowledge and interviewer evaluations present a more heterogenous distribution. What is the implication of this result for the rest of the estimations? This needs to be discussed.

With respect to the validation of the measure in Figure 3, it is true that the estimates corresponding to discursive sophistication appear to be of a greater size than those corresponding to factual knowledge. However, it is also true that the precision of the estimates is also smaller (this is especially relevant for 2018 CES and 2016 ANES estimations)

Gender differences Figure 6 shows that average levels of discursive sophistication are systematically smaller than factual knowledge. This suggests that discursive sophistication is a quality that only those highly motivated (and perhaps partisan) citizens have. So, we have the situation of low average level of discursive sophistication with no gender differences. I wonder if this is the same for the other sources of inequality in political sophistication (such as education or age) that the extant literature has found.

Reviewer 3

I've read this paper carefully. The text analysis methods and the structural topic model fall outside my area of expertise. However, I consider myself methodologically sophisticated and familiar with the literature on political knowledge and the gender gap.

Here's my take on the paper: It's wonderful. I have a couple of minor, easy-to-implement suggestions that I suspect will marginally improve the paper. But it's wonderful as-is. My formal recommendation is minor revisions.

The paper is well-written. It has a compelling conceptual argument and a compelling empirical argument. It discusses a theoretically and normatively interesting topic. I can't remember ever reviewing a paper I enjoyed this much—I like everything about it. I regret only that I didn't complete it sooner, because I would have loved to let my undergraduates read the paper this semester.

The Comparability of the Effects of Discursive and Factual Knowledge

I would like to see the concerns below addressed in a revised submission. I suspected they will require only a little work, and that any changes will improve (and not undermine) the argument.

One small suggestion, "average marginal effect" is a bit vague. Perhaps I missed it, but I want to confirm that the effects shown in Figure 3 are comparable across discursive and factual knowledge. For example, are these the change in the probability of voting as the individual moves from the 25th to 75 percentile on each measure? If the factual and discursive measures are scaled differently, then the results are not comparable. It doesn't seem trivial to select a comparable shift for the two measures. Even if both are scaled from 0 to 1, then the two effects might not be comparable, because a shift from .3 to .5 might be a shift from a person with very low discursive knowledge (say, 20th percentile) to very high (say, 80th percentile), while a comparable shift in factual knowledge might

be 0.2 to 1.0. See the different distributions of factual and discursive knowledge in Figure 2, for example.

Alternatively, the authors might make the measure of discursive and factual knowledge comparable by rescaling the SD of each to one. This does not guarantee a comparable effect, but I think it's a much more plausible default than rescaling from zero to one.

The same argument applies to Figure 4. It could be that the factual measure has many observations along the entire range from 0 to 1, while almost of the discursive measures fall between .2 and .5. In this case, increasing both measures from there 20th to 80th percentiles (a comparable shift, I suggest) would produce a similar difference in retrieval.

It also isn't clear why the authors use control variables in these models. I don't understand the purpose of the control variables (not causal inference, see middle paragraph on p. 17), so it's difficult to evaluate whether these are the appropriate control variables or not. The context makes me think the authors might want no control variables at all—just a simple linear or logistic regression with factual/discursive knowledge as the explanatory variable.

Rather than look at effect size at all, the authors might consider comparing the fit of the two models against each other using the AIC and/or BIC. I suspect this is the route I would have chosen if I had been trying to make a similar point.

Arguments for the Null Effect

This is merely a suggestion for the authors. If they find my suggestion unhelpful or misguided, they should feel free to ignore it.

I also suggest that the authors think more carefully about their argument for no gender gap. Rainey (2014, AJPS, "Arguing for a Negligible Effect") shows that a lack of statistical significance is neither necessary nor sufficient to demonstrate "no effect." Instead, he argues that the researcher should/must argue that all the effects in the 90% confidence interval are substantively negligible. Looking at the CIs in Figure 7, it seems that the CIs contain only very small effects, but perhaps the authors should make the explicit argument that these effects are indeed negligible.

Related to the point above, it is critical that the effects shown in Figure 7 are comparable. I suspect it's not the case, but it's possible that a 0.01 gap in discursive knowledge is "large," while a 0.05 gap in factual knowledge is "small." It might be easiest to think of effect sizes in terms of standard deviations or percentile shifts in the original outcome. For example, perhaps a 0.01 gap in discursive knowledge is a 1 SD shift or a shift from the 25th to the 75th percentile, while a 0.05 shift in factual knowledge is a 0.25 SD shift or a shift from the 45th to the 55th percentile. (Figure 21 shows the distributions, so I can almost work out—but not quite—whether these effects are comparable.)

Figure 6

This is merely a suggestion for the authors. If they find my suggestion unhelpful or misguided, they should feel free to ignore it.

Figure 6 is awesome, and it makes a really powerful point. However, I wonder if a histogram of the entire distribution (for men and women separately, rather than showing only the average for each) would be informative. A “violin plot” or a “beeswarm” plot could show the entire distribution for men and women in a similar space. It would use the space above the error band (currently empty) and the space below the error band (filled by a not-strictly necessary bar) to show the entire distribution.

It could be that showing the entire distribution for men and women is not helpful, but I’m curious, as a reader, what it looks like.