Response to Reviewers

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

The following memo details revisions I have made 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 their thoughtful suggestions. I have addressed each of their comments directly in the revised manuscript and/or by providing additional analyses in the appendix. For example, I have 1) revised the framing to emphasize that factual knowledge and discursive sophistication can be viewed as complements rather than competing alternative measures, 2) added a theoretical discussion of the role of declarative vs. procedural memory for political competence, and 3) conducted supplementary analyses to provide more information about the antecedents of discursive sophistication and to show that it captures a distinctively political *rather than a more* general *phenomenon. Lastly, I have implemented suggested improvements of the statistical analyses reported throughout the manuscript as well as the corresponding visualizations. I am very grateful for the useful and insightful comments provided by all three reviewers and I believe that each of them have helped me produce a stronger revised manuscript that clarifies important theoretical arguments and key empirical results.*

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.

Thank you for this concise summary of our paper. I will directly address each concern below.

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[1] that the two are only modestly correlated, and in Figure 3[2] that both produce effects on typical political dependent variables (although it does not appear that both variables are include 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.

This is a good point and I have revised the introduction, theory, and results sections accordingly. Specifically, I now emphasize throughout the manuscript that both measures should not be viewed as competing alternatives. On page 2, for example, I now write: "While discursive sophistication shares a considerable amount of variance with traditional metrics, they are far from equivalent. Indeed, discursive sophistication and factual knowledge are independent predictors of turnout, political engagement, and various manifestations of political competence—suggesting that both measures can be viewed as complements that capture different aspects of political sophistication." Furthermore, I have changed the model specifications underlying Figures 2 and 3 such that both variables are included as IVs at the same time (see Tables D.1–D.5) and revised the corresponding discussion of results to reduce the appearance of a competition between both measures and to emphasize how both contribute to our understanding of political attitudes and behavior.

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.

I thank R1 for this thoughtful suggestion. I have now included additional models incorporating the

---

[1]*Figure 1 in revised manuscript.*
[2]*Figure 2 in revised manuscript.*

suggested interaction effects in the appendix tables for Figures 2 and 3 (see Tables D.1–D.5). Interestingly, these models reveal positive and statistically significant main effects for both measures while the interaction coefficients are largely null. There are a few exceptions, however, where we additionally observe positive *interactions between discursive sophistication and factual knowledge, which suggests that both concepts are mutually reinforcing. I have included a brief discussion of these findings in Footnote 15. However, these patterns do not differ by gender, which is why I have not included these additional results (i.e., interaction effects broken up by gender) in the current manuscript (although I'd be happy to add them to the appendix as well if R1 prefers).*

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 and important general aspect of sophistication, whereas civics knowledge contributes a political dimension.

This is another important point raised by R1. I have revised the theory section and added a more elaborate treatment of this issue in the discussion section. On p. 12, for example, I describe how we can "study discursive sophistication in well-defined (and potentially narrow) areas by using open-ended questions that focus on attitudes and beliefs that are relevant for a specific context. Researchers interested in citizen competence in local politics, for instance, could field a battery of open-ended questions examining relevant topics such as schooling, zoning, or other areas of local administration." The idea here is that selecting open-ended items with a certain topical focus would allow researchers to capture discursive sophistication in well-specified areas of interest. I return to this idea in the discussion on pages 34-35 to indicate that more work is necessary to empirically test this argument, for example by "comparing discursive sophistication based on open-ended responses centered around politics with an equivalent measure based on non-political questions such as sports, literature, or science. Going forward, this line of research could develop

best practices for the selection of different question types (e.g., targeting specific policies vs. party evaluations) to measure discursive sophistication."

In addition to these theoretical considerations, I have included the supplementary analyses suggested by R1 in Appendix C.VI. Specifically, Table C.2 shows that discursive sophistication and factual knowledge has diverging associations with certain personality characteristics, verbal skills, and survey mode.[3] For instance, while openness to experience has a positive effect on discursive sophistication, it has a negative effect on factual knowledge (at least in the 2012 ANES). Being reserved, on the other hand, shows a negative association with discursive sophistication but no relationship with factual knowledge. Especially interesting, however, is the finding that verbal skills (measured using the Wordsum vocabulary test) have a stronger effect on factual knowledge than discursive sophistication. Furthermore, Respondents in online surveys score significantly higher on factual knowledge than in face-to-face interviews. This difference can be attributed to the fact that individuals are able to look up answers for factual knowledge questions while taking an online survey (Clifford and Jerit, 2016). For discursive sophistication, on the other hand, individuals perform better in the face-to-face survey. Open-ended answers in online surveys may be less elaborate because respondents have to manually type their responses. These results illustrate once again that both measures should be seen as complements rather than competing metrics of political sophistication, as they capture different aspects of the underlying concept of interest.

Regarding the effect of differentiated education categories, Table C.2 indicates that higher education categories (e.g., college degrees) have stronger effects on discursive sophistication than lower levels of education (e.g., high school degrees). The same is the case, however, when looking at the association between these education categories and factual knowledge. The only suggestive evidence for divergence between both measures in this context is the finding that high school education appears to have a more consistent positive effect on factual knowledge while its effect on discursive sophistication does not reach conventional levels of statistical significance—which suggests that discursive sophistication is particularly fostered among higher levels of education. Even after controlling for these additional predictors, factual knowledge remains significantly correlated with discursive sophistication, which suggest that both share an underlying latent construct that cannot be reduced to any of the potential confounding factors mentioned herein.

Lastly, I have now conducted additional analyses in Appendix C.IV and C.V that further illustrate how discursive sophistication is distinctly related to political *competence. Specifically, respondents who score higher on discursive sophistication display a smaller degree of uncertainty around the ideological placement of politicians and parties (Figures C.4 and C.5)—and are ultimately more likely to vote based on ideological proximity in senatorial races (Figure C.6).*

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. Chang-

---

[3]These analyses are based on the 2012 and 2016 ANES, where additional measures of personality, verbal skills, and survey mode were available.

ing the title would help with decoupling discursive sophistication from knowledge, thereby improving the paper's framing.

I have revised the framing of the paper (see discussion above) but followed the recommendation of the editors to keep the title as-is.

---

2. Luskin's 1990 Political Behavior paper is worth a look.

Thank you for this suggestion, I added a reference to this paper in the theory section (p. 4).

---

3. It is Delli Carpini, not Carpini. "Delli" is part of the last name, not a middle name.

Fixed.

---

4. Figure 1 is not necessary.

Agreed. I have removed Figure 1.

---

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).

That is accurate, I added the corresponding reference in the theory section on page 5.

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 knowledge 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:

Thank you for these supportive comments, I will respond to each of R2's specific points below.

---

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)

These are great suggestions, I have revised theory section and literature review to discuss the role of declarative vs. procedural memory as well as the role of question format. For instance, on page 4 of the revised manuscript, I have added references to Lupia (1994); Lau and Redlawsk (2001) and Lau, Andersen, and Redlawsk (2008) highlighting the role of heuristics as a potential substitute for declarative memory. I also discuss Prior and Lupia (2008) and Bernhard and Freeder (2020) to assess whether political learning skills can facilitate informed decisions when declarative memory is low. Furthermore I included a more elaborate discussion of the potential impact of question format and added a reference to Fraile and Fortin-Rittberger (2020) and other related research (see for example p. 5 of the revised manuscript).

---

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.

Agreed, I have added these references and a brief discussion to my review of previous research on the gender gap (see pages 5–7 in the revised manuscript).

---

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.

This is a point well taken. I have made revisions throughout the manuscript to highlight theoretical contributions to the literature (e.g., by discussing the connection between the structure of belief systems and political competence in more detail), although I agree with R2 that the core contribution of the manuscript is more methodological and empirical. In other words, this manuscript builds on the seminal concept of the structure of belief systems—which is still influential but is not reflected in conventional knowledge metrics—and develops a theoretically grounded measurement framework using automated text-as-data methods rather than having to rely on manual coding (which is presumably why the concept has not been used as much in recent research). To a certain extent, the lack of additional theoretical contributions is driven by the fact that developing a new measurement approach for an existing theoretical construct such as political sophistication necessitates a solid foundation in the established literature. Proposing a completely new measure and *a completely new theory at the same time risks obfuscating the connection to existing work.*

---

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.

*I agree with R2 and have made additions in the measurement section that explicitly discuss this point (see p. 8-9 & 12-13). For example, I now write on page 9: "To what extent does political sophistication defined as a complex system of beliefs ultimately facilitate citizen competence? As discussed above, conventional knowledge questions have been criticized because the required information has little relevance for people's ability to make high-quality decisions (*Lupia*, 2006). As *Cramer and Toff* eloquently summarize, conventional measures implicitly focus on "what people do not know" (2017, 756, emphasis added) by presupposing pieces of information as necessary for political competence. Examining people's political beliefs system, on the other hand, allows us to shift the focus back to what they do know and how they use that information. After all, a large, wide-ranging, and highly constrained system of beliefs will help citizens to locate their own interests within the political system, understand the functioning of institutions, assess the performance of the incumbent government, and evaluate the actions of the main political actors (e.g., *Converse*, 1964)." Furthermore, I included supplementary analyses in the appendix that address the question empirically. Specifically, Appendix *C.IV* and *C.V* show that discursive sophistication is distinctly related to political competence in the sense that respondents who score higher on discursive sophistication display a smaller degree of uncertainty around the ideological placement of politicians and parties (Figures *C.4* and *C.5*)—and are ultimately more likely to vote based on ideological proximity in senatorial races (Figure *C.6*).*

---

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?

I thank R2 for raising this question, since it is crucial to address this point in order to develop best practices for applied researchers who want to leverage open-ended responses in the future. It is worth noting that the surveys presented throughout the manuscript rely on different sets of open-ended questions that vary with regard to their level of partisan polarization. On one end of the spectrum, we employ highly salient policy questions (focused on gun legislation, abortion, etc.) in the CES and YouGov study, which could be viewed by many as wedge issues. The open-ended likes/dislikes questions employed in the ANES, on the other hand, do not necessarily invokes these types of issues, but rather asks respondents to explain their preferences for one party or candidate over another (based on whatever consideration they find most salient). On the other end of the spectrum, we have the open ended items employed in the Swiss survey which focus on the content of various policy referenda, many of which are less controversial and do not overshadowed by partisan polarization. Thus, our results hold while covering a range of open-ended questions that vary in partisan content.

This leaves open the question whether applied researchers should focus on more or less partisan content in their open-ended responses. In the revised measurement section, I emphasize that from a theoretical perspective, the guiding principle should be to focus on issues that are relevant for political decision-making. For instance, if we are interested in whether make well-informed vote choice, then we should measure discursive sophistication in the context of issues that voters deem relevant for their vote. This might imply, of course, focusing on polarizing partisan content, since these are the issues people care about when choosing who to vote for (take the issue of abortion in the context of the upcoming US midterms as an example).

Another reason why avoiding partisan content may be counterproductive is related to the earlier concern by R1 raising the issue whether discursive sophistication is capturing something that is uniquely political or rather more general communication skills. While I argue in my response to R1 that discursive sophistication is indeed specific to politics, deviating too far from salient issues that are relevant for political decision-making may increases the risk of potential confounding due to verbosity etc.

In the revised manuscript, I have added some of the arguments outlined here in the measurement section for discursive sophistication. Furthermore, I have now included a statement in the discussion section that more work is necessary to develop best practices when it comes to open-ended item selection to measure discursive sophistication. Lastly, I have included partisanship as a predictor of discursive sophistication and factual knowledge in the supplemental analyses reported in Table C.2. Interestingly, the evidence suggests that there are no significant partisan differences in discursive sophistication (thereby alleviating potential concerns about partisan content). If anything, there is evidence that factual knowledge is associated with partisanship.

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.

As R2 points out, the main difference between these measures is that discursive sophistication is truly continuous (and its distribution is close to normal), whereas factual knowledge (and interviewer evaluations displayed in Figure 1) are technically ordinal but treated as "quasi-continuous." Regarding the consequences of this difference for the remaining estimations, two related considerations need to be raised. On the one hand, using a truly continuous measure adds variation, which should result in more precise estimates if it is used as an IV (all else being equal) and may increase (or decrease) model fit if it is used as a DV (depending on the underlying covariance structure). On the other hand, we have to consider the impact of two potential sources of measurement error: 1) estimation uncertainty inherent to discursive sophistication (due to modeling assumptions when processing text as data etc.) and 2) forced discretization inherent to additive knowledge scales (due to measuring a continuous latent construct using a discrete scale). Depending on which of these sources of measurement error is larger, we may see more uncertainty and/or attenuation bias for one metric or the other. While quantifying and comparing these different sources of measurement error is outside of the scope of this article, it is an important issue that I hope to investigate in future studies. For now, I have added a brief discussion of these considerations in Footnote 14.

With respect to the validation of the measure in Figure 3[4], 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)

This is an important observation. Part of the reason why the estimates corresponding to discursive sophistication in the original version of Figure 2 appear larger while at the same time exhibiting greater estimation uncertainty is the fact that although both measures were rescaled to range from 0 to 1, the underlying distributions differed (with factual knowledge exhibiting a larger standard deviation than discursive sophistication; see R3 comments and response for details). The original figure displayed average marginal effects for a unit change in each independent variable, which implicitly resulted in an unequal comparison with regard to the underlying percentiles in discursive sophistication and factual knowledge. In response to the comments by R3 outlined below, I have rescaled both discursive sophistication and factual knowledge to zero mean and unit variance / standard deviation, which results in a more appropriate comparison of effect sizes in Figure 2. The revised plot shows that discursive sophistication and factual knowledge have effects of similar magnitude and precision. Keep in mind, however, that the underlying model specifications were further revised in response to R1's previous suggestions and now include both measures simultaneously— although this change had no systematic impact on the relative effect size or estimation uncertainty.

---

[4]*Figure 2 in revised manuscript.*

Gender differences Figure 6[5] 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.

Similar to the previous point, this is issue has been fixed by rescaling both measures to zero mean and unit variance. The revised version of Figure 5 now shows no sign of systematic differences between average levels of discursive sophistication and average levels of factual knowledge. Since this equivalence across measures is now essentially constructed by design (since both metrics were mean-centered), it is important to elaborate on why this step is appropriate.

First, recall that both measures target the same underlying latent construct (i.e., political sophistication, albeit different components of it). Given that both metrics rely on fundamentally different approaches, there is no reason to expect perfect measurement invariance and/or scaling equivalence between them. In other words, both metrics represent related scales that—despite having been rescaled to range from 0 to 1 in the original submission—ultimately have diverging underlying distributions (i.e., different means and variances). One reason why these resulting distributions may diverge is the observation that conventional political knowledge scores can suffer from ceiling effects (i.e., lack of differentiation between respondents who answer all knowledge questions correctly, see added discussion in the revised manuscript on p. 20). These ceiling effects imply that the maximum value of factual knowledge subsumes a potentially large range of values in the underlying latent variable, which in turn explains higher average values in factual knowledge compared to discursive sophistication (if both are scaled from 0 to 1). Now, given that both metrics ultimately capture the same latent variable (or closely related subcomponents) with different measurement properties, it is more appropriate to compare metrics that hold constant the basic features of the underlying distribution (i.e., same mean & variance) rather than to hold constant the end-points of both distributions (i.e., scaling from 0 to 1)—particularly if one of the measures under consideration suffers from ceiling effects.

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.

---

[5]*Figure 5 in revised manuscript.*

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.

I thank R3 very much for their support and these encouraging comments.

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[6] 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[7], 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.

This is a great point. In the original submission, both discursive sophistication and factual knowledge were scaled to range from zero to one but the underlying distributions are indeed different (with factual knowledge exhibiting a larger standard deviation than discursive sophistication). As R3 correctly points out, this makes it difficult to compare average marginal effect sizes, which illustrate the expected change in each outcome measure for a unit increase in each predictor.

I followed R3's suggestion to address this issue by rescaling both measures to zero mean and unit variance / standard deviation and clarified the type of comparison in Figure 2 (now showing the expected change in each outcome for an increase in each predictor from 1 SD below the mean to 1 SD above the mean). In addition, I changed the example in Table 2 to directly illustrate the

---

[6]*Figure 2 in revised manuscript.*
[7]*Figure 1 in revised manuscript.*

substantive meaning of such a change in discursive sophistication. Although the relative effect sizes of discursive sophistication compared to factual knowledge in Figure 2 changed slightly, the substantive conclusions remain the same: both measures are significant predictors of important outcomes that are usually viewed as associated with political sophistication. This holds despite the fact that I now included both measures in a single model, which suggests that both are complements rather than competing indicators (see also the previous discussion in response to R1 on this point).

These revised results are equivalent to computing marginal effects for changes between specific percentiles as initially suggested by R3 above. I have decided to directly plot the expected effect for SD changes rather than specific percentiles, since rescaling both measures additionally addresses R3's later point regarding the comparability of gender differences in discursive sophistication and factual knowledge (see further discussion below).

The same argument applies to Figure 4[8]. 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.

This issue has been fixed by converting both knowledge measures to zero mean and unit variance (see discussion above). I have revised Figure 3 accordingly and adapted the corresponding discussion in the manuscript. I believe that these changes strengthen the manuscript and provide additional insights, since the comparison of the rescaled measures reveals how conventional additive knowledge scales can suffer from ceiling effects since there is no way to differentiate respondents who answer all questions correctly (or incorrectly, although that is less common with standard batteries). Discursive sophistication suffers from no such constraints and therefore allows us to better represent the full spectrum of the underlying latent variable.

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.

I decided to use sociodemographic controls in order to address potential concerns regarding confounding factors (e.g., due to varying levels of education etc.). That said, I have now included results for basic models without controls in the Appendix (see Tables D.1 through D.5. The substantive results regarding the comparison of discursive sophistication and factual knowledge remain unchanged across all models when omitting control variables.

---

[8]*Figure 3 in revised manuscript.*

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.

As discussed in the response to R1, I have now reframed the paper to reduce the appearance of a competition between factual knowledge and discursive sophistication as mutually exclusive alternative measures. To that end, I have also revised the models to include both measures simultaneously in each model specification. Given these changes, the comparison of model fit is not directly applicable since the goal is now to show that both are complementary measures.

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[9], 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.

Thank you for this suggestion, I have added a brief discussion of (Rainey, 2014) and included the suggested tests for negligible effects in Figure 6. As described by R3, these revisions consist of displaying 90% confidence intervals around the estimates in Figure 6 and showing that these CIs contain only small effects. Rescaling discursive sophistication and factual knowledge proved useful here as well, since it allows for a direct identification of small effect sizes (equivalent to Cohen's d ≤ 0.2; see Sawilowsky, 2009). Revised Figure 6 reveals only negligible differences in discursive sophistication between women and men across, whereas the gender gap in factual political knowledge remains substantively an statistically significant.

Related to the point above, it is critical that the effects shown in Figure 7[10] 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.

---

[9]*Figure 6 in revised manuscript.*

[10]*Figure 6 in revised manuscript.*

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 2[11] shows the distributions, so I can almost work out—but not quite—whether these effects are comparable.)

Agreed, I addressed this point by rescaling both measures to unit variance / standard deviation (discussed above) and revising Figure 6 as well as the corresponding discussion in the manuscript. Gender differences are now shown in standard deviations rather than on a 0 to 1 scale, which facilitates the comparison between both measures. The substantive results remain the same, we only observe negligible gender differences for discursive sophistication and sizable differences for factual knowledge (in SD units).

Figure 6[12]

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[13] 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.

This is another excellent suggestion. I have replaced the bar chart in Figure 5 with a beeswarm plot that allows for a direct comparison of the respective distributions. Furthermore, I included visual cues to directly assess the statistical significance of the gender differences across all surveys. I believe that this revised figure provides a lot more useful information than the previous version and I really appreciate R3's comment on this point.

---

[11]*Figure 1 in revised manuscript.*
[12]*Figure 5 in revised manuscript.*
[13]*Figure 5 in revised manuscript.*