

Let’s Talk Politics

A Naive Approach for Measuring Political Sophistication

Patrick W. Kraft

**Abstract**

This paper proposes a simple but powerful framework to measure political sophistication based on open-ended survey responses. *Discursive sophistication* utilizes automated text analysis methods to capture the complexity of individual attitude expression. I validate the approach by comparing it to conventional political knowledge metrics in multiple studies using different batteries of open-ended items. The paper then illustrates how the measure can help refine previous insights from the literature such as 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 beliefs.

Factual knowledge scores are problematic because they...

- ... ignore partial knowledge (e.g. DeBell, 2013)
- ... are biased due to guessing (e.g. Mondak and Anderson, 2004)
- ... convolute different types of knowledge (Barabas et al., 2014)
- ... do not capture belief systems (Luskin, 1987; Tetlock, 1983)
- ... are not related to political competence (Lupia, 2006)

Discursive sophistication captures the complexity in open-ended responses

Facing a set of questions on various political issues, sophisticated respondents should be able to express a larger number of considerations, use words that are highly descriptive of each topic, and be able to voice their opinion on each issue.

$$\text{considerations}_i = \frac{|\mathcal{T}_i^*|}{\max_{i'} |\mathcal{T}_{i'}^*|}$$
$$\text{word choice}_i = \frac{\sum_{\mathcal{W}_i} P(w|t^*)}{\max_{i'} \left[ \sum_{\mathcal{W}_{i'}} P(w|t^*) \right]}$$
$$\text{opinionation}_i = \frac{-\sum_{j=1}^J p_{ij} \ln p_{ij}}{\ln J}$$
$$\text{sophistication}_i = \frac{1}{3}(\text{considerations}_i + \text{word choice}_i + \text{opinionation}_i)$$

$i$ Individual respondent

$\mathcal{T}_i^*$ Set of topics mentioned by respondent  $i$  (via STM, see Roberts et al. 2014)

$\mathcal{W}_i$ Complete set of words contained in all responses of person  $i$

$w \in \mathcal{W}_i$ Individual word in a response, each assigned to a topic  $t^* \in \{1, ..., T\}$

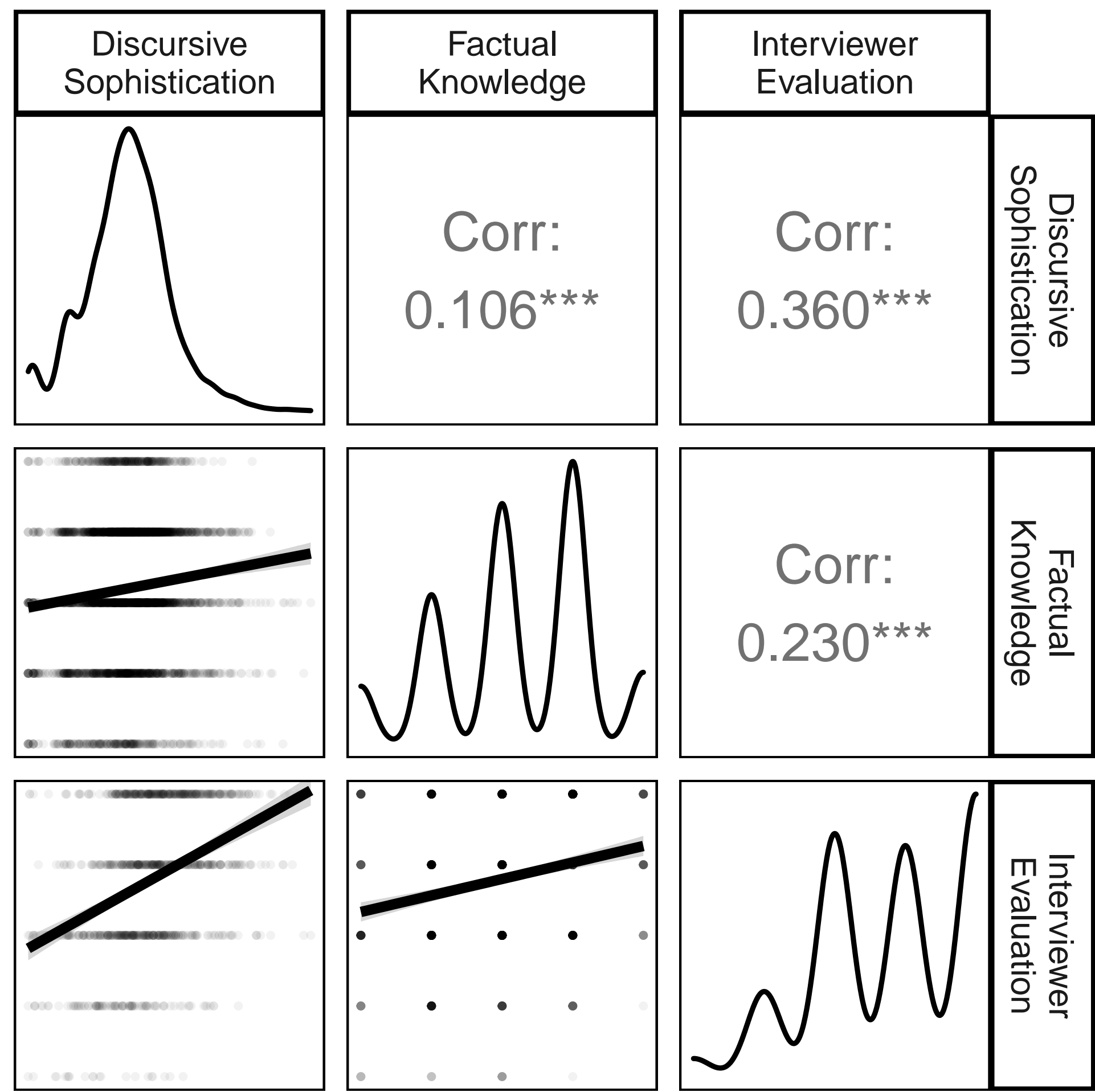
$p_{ij}$ Proportion of words in person  $i$ 's response to question  $j \in \{1, ..., J\}$  relative to the overall response length across all questions.

Consider the following illustrative example

Low Sophistication Response		High Sophistication Response	
Clinton (+)		Politician.	
Clinton (-)	The fact that she has links to Al-Qaeda.	Caught in lies.	
Trump (+)		Says what he thinks.	
Trump (-)	He is going to start a civil war. I feel like he is racist.	Reality TV star, poor business-man	
Democrats (+)		Middle class minded.	
Democrats (-)		Too many handouts.	
Republicans (+)		Economic growth conscious.	
Republicans (-)		For the big business.	
Disc. Soph.	0.162	0.461	

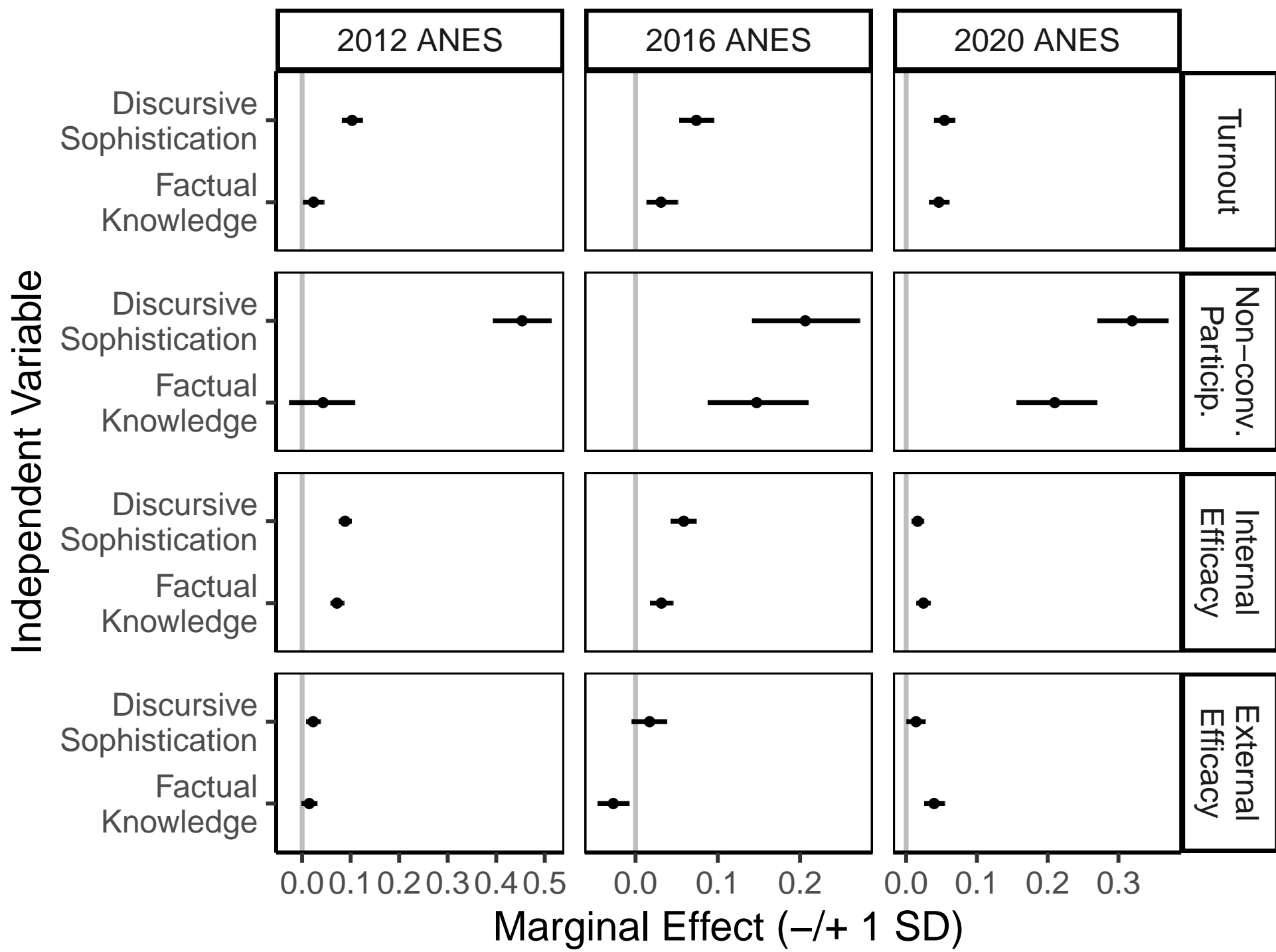
**Table 1:** Example of open-ended responses for low and high scores on discursive sophistication with equal factual knowledge scores (3 out of 4 correct responses) in the 2016 ANES.

Discursive sophistication is correlated with conventional knowledge measures...



**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 are statistically significant with  $p < .05$ .

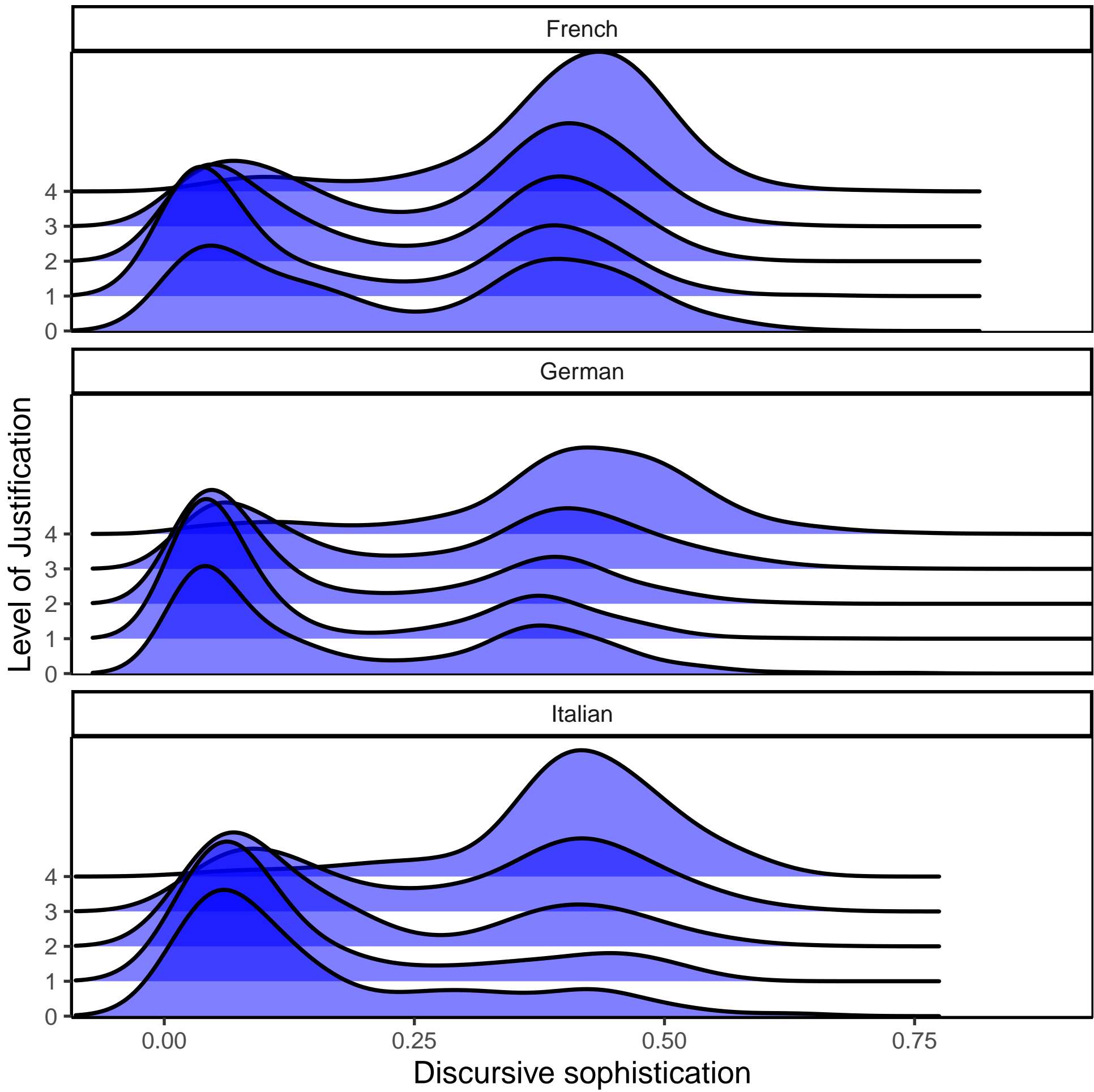
...but it is a better predictor of engagement and participation in politics...



**Figure 2:** Effects of sophistication on turnout, non-conventional participation, internal efficacy, and external efficacy in the 2012 and 2016 ANES. For each dependent variable, the figure displays the change in expected values after increasing each sophistication measure from -1 to +1 standard deviation from its mean (including 95% confidence intervals). Model estimates are based on logistic regression (turnout) or OLS (non-conventional participation, internal efficacy, external efficacy). Both sophistication measures are included simultaneously while controlling for gender, education, income, age, race, church attendance, survey mode, and Wordsum vocabulary scores.

University of Wisconsin-Milwaukee  
Department of Political Science  
✉ kraftp@uwm.edu  
🐦 @patrickwkraft  
🌐 pwkraft.github.io

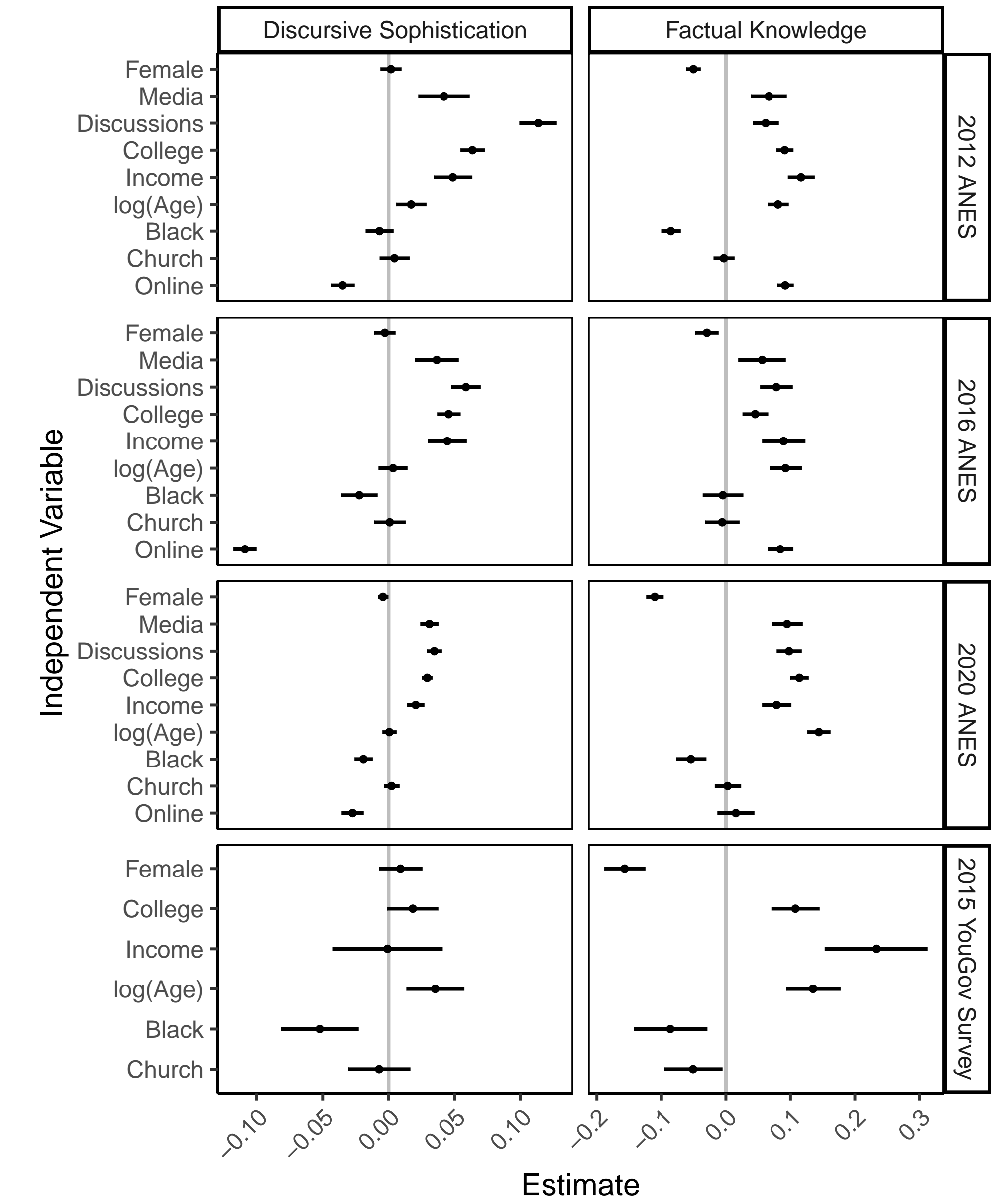
...and it is positively related to manual coding of well-justified preferences



**Figure 3:** Discursive sophistication and manually coded level of justification (Colombo, 2016) in Swiss post-referendum surveys. The plot compares kernel densities of discursive sophistication for each manually coded level of justification.



Application: No evidence for a gender gap in discursive sophistication



**Figure 4:** Common determinants of political sophistication. Estimates are OLS regression coefficients with 95% confidence intervals. Dependent variables are discursive sophistication as well as factual political knowledge.



Why no gender gap? Women can focus on different issues in open-ended responses

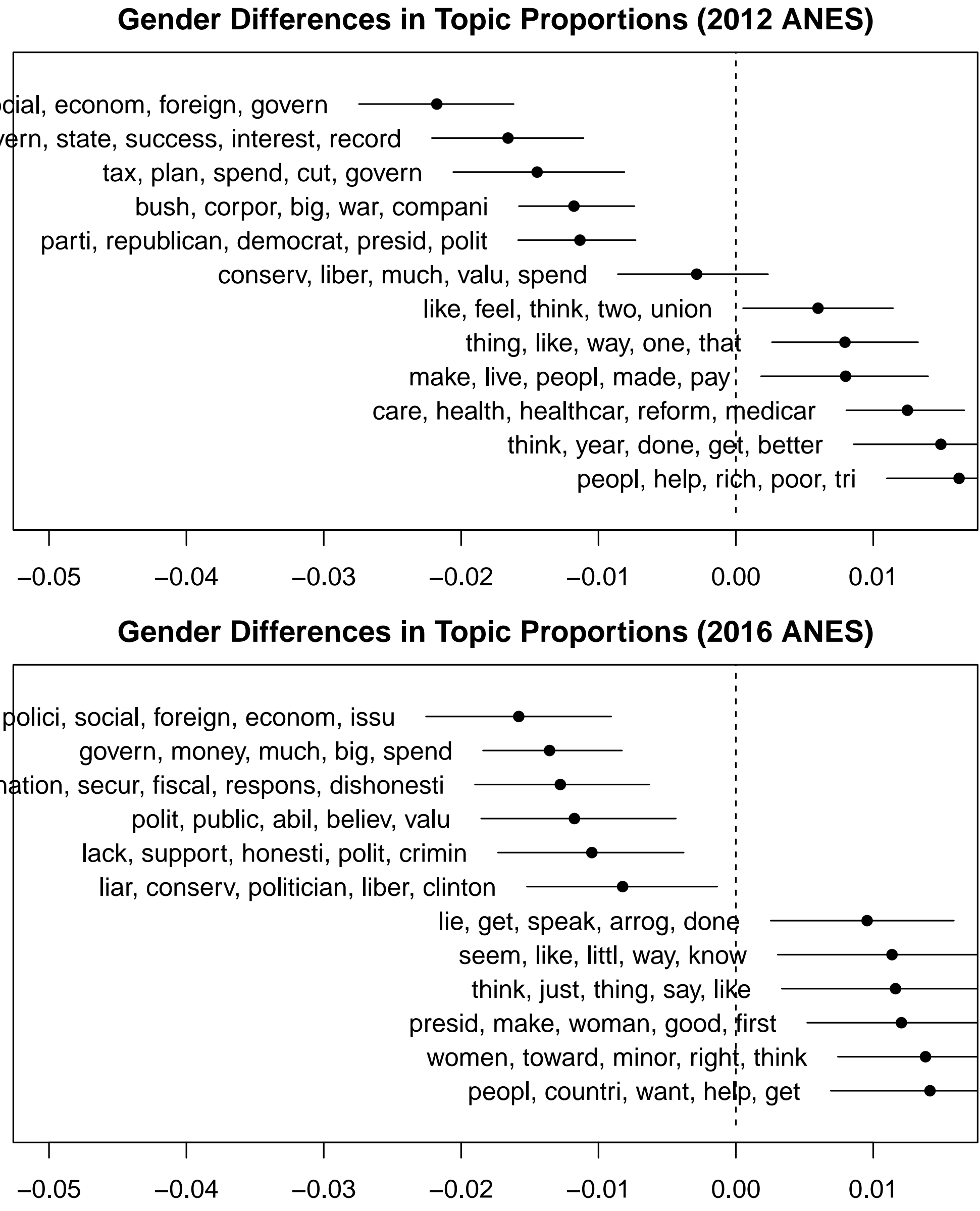


Figure 5: Gender differences in topic proportions in open-ended responses based on the structural topic model used to compute discursive sophistication (including 95% confidence intervals). Coefficients indicate the difference in predicted topic prevalence among men and women; positive values indicate higher prevalence among women. Labels are based on the five highest probability terms related to the topic.

Would you like to know more?  
Download the paper here:



References

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

Colombo, Céline. 2016. "Justifications and Citizen Competence in Direct Democracy: A Multilevel Analysis." *British Journal of Political Science*: 1–20.

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

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

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

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.

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.

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