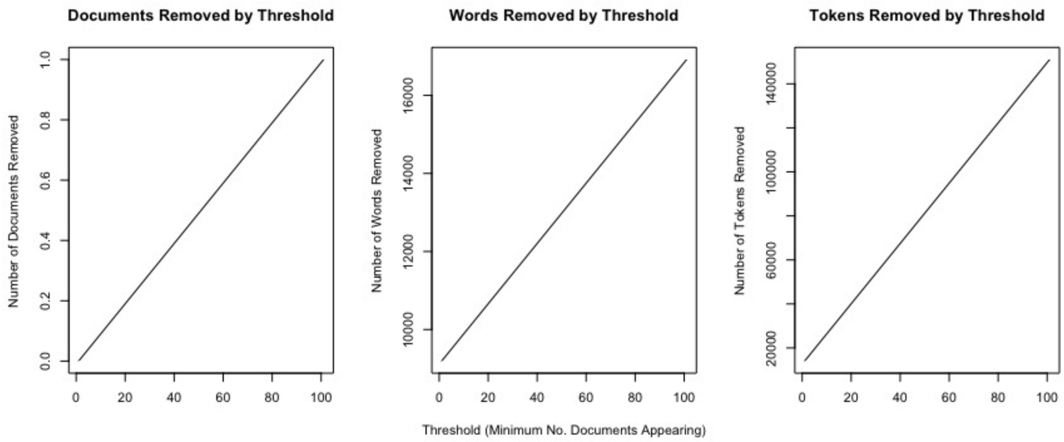
A Appendix

A.1 Pre-Processing & Topic Selection Strategy

Figure A.1: Document, Word, and Token Removal as a function of Threshold

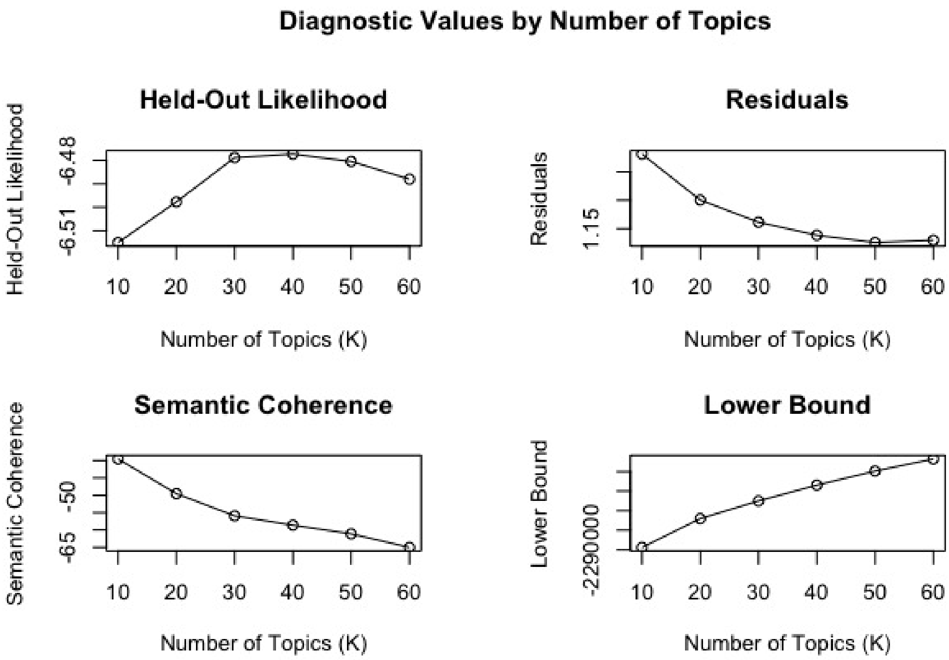


Results were replicated using several different thresholds: 15, 20, and 25. Topics produced and their word associations were nearly identical to those presented in the body of the paper.

We employ a document removal threshold of 30, meaning words that do not appear in at least 30 documents are dropped for computational efficiency. We assume that words appear in such a few number of documents will offer little information for our analysis. Modifying the document removal threshold threshold to 25, 20, and 15 produced substantively identical results in our subsequent analysis. After removing stop words, punctuation, numbers, and candidate names as well as stemming words to their root, we are left with a corpus of 1,507 documents and 217,536 tokens.

To run our initial STM we set the model initialization to `"Spectral,"` which uses the connection of LDA with non-negative matrix factorization that provides theoretical guarantees that the optimal parameters will be recovered. We set the number of topics, K, to zero, which automatically uses an algorithm developed by Mimno et al. (2011) to select the number of topics. Roberts et al. (2014a) stress that this approach does not select the “true” number of topics but rather is a useful place to start. This initial model produced a total of 48 topics.

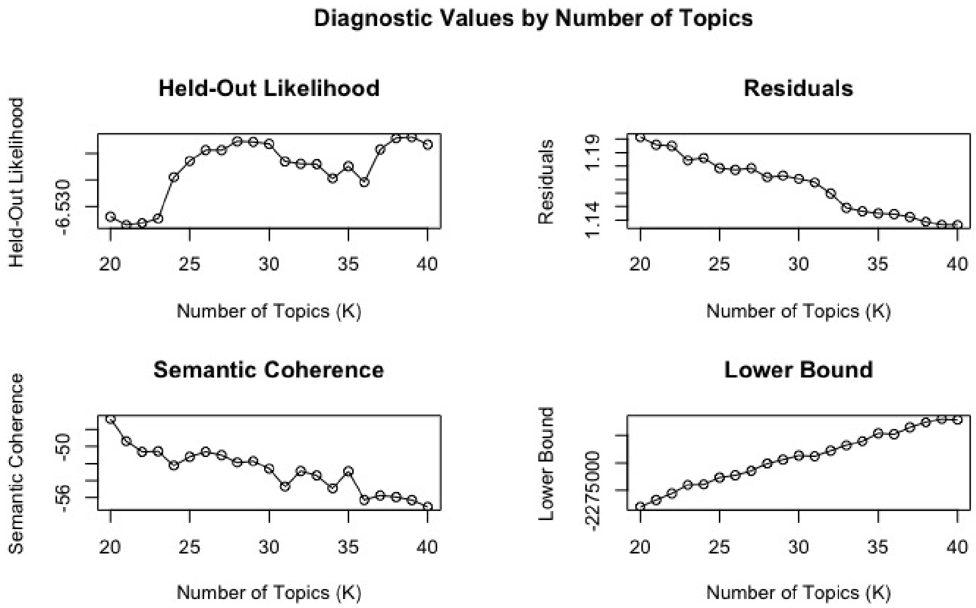
Figure A.2: Diagnostic Values for STM Model (Sequence of Topics 10-60)



Model value K varies across the sequence: 10, 20, 30, 40, 50, 60. Model specified with factor covariate from Table A.1 and successfully converged

After finding an initial number of topics, we ran several STMs, varying the number of topics across a sequence of values from 10 to 60 using the `searchK` function in the `stm` package. The results are presented in Figure A.2. Four metrics are provided to assess the model quality: held-out likelihood, residual dispersion, topic semantic coherence and the approximation to the lower-bound of the marginal likelihood. The goal is maximize the held-out likelihood, semantic coherence, and lower-bound while minimizing the dispersion of residuals. Referencing Figure A.2, we ran two more `searchK` functions narrowing our sequence of topics each time based on metric performance. Our final run included a sequence of K topics from 20 to 40. The results are presented in Figure A.3. We settled on specifying our model with 27 topics. Modifying the number of topics slightly, increasing and decreasing K by one produced results nearly identical to those presented in the body of the paper.

Figure A.3: Diagnostic Values for STM Model (Sequence of Topics 20-40)



Model value K varies across the sequence of integers 20 through 30. Model specified with factor covariate from Table A.1 and successfully converged.

To determine the strength of our modeled topics we produced a plot of topic quality, which is displayed in Figure A.4. Topic quality is evaluated using semantic coherence — how often words within a topic co-occur — and exclusivity — the uniqueness of words to each topic — which are displayed on the x-axis and y-axis respectively. Using this approach is accepted as a reasonable surrogate for human judgment on the quality of topics (Mimno et al., 2011). The highest quality topics fall in the top right corner of Figure A.4. Based on the semantic coherence and exclusivity metrics, around six topic produced by our model, including *Nature/Land, Immigration, Agriculture, and Partisan Issues* could be considered "lower-quality." However, reviewing topic prevalences in Table 1, these lower quality topics are also those that occur less often in the text, which lessens concern that our interpretation of results is not meaningful.

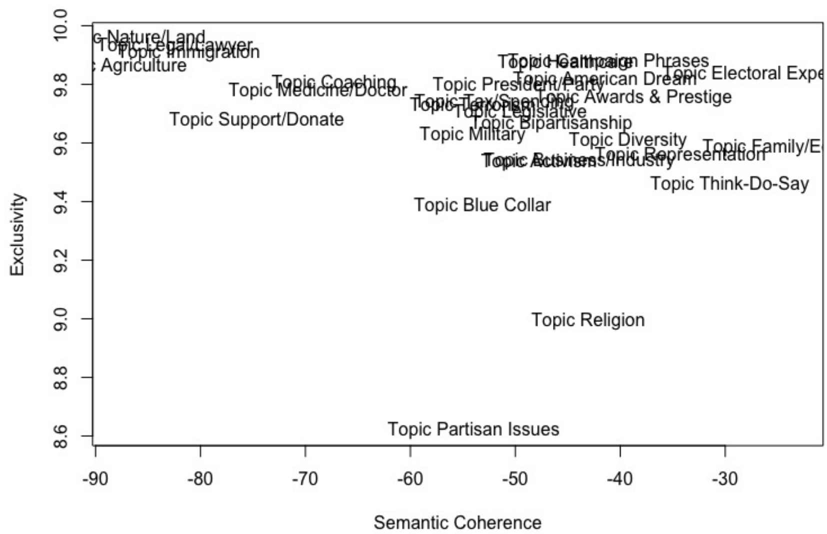


Figure A.4: Partisan-Gender-Experience Model Topic Semantic Coherence and Exclusivity

Note: Semantic coherence, on the x-axis, refers to how often words within a topic co-occur. Exclusivity, on the y-axis, refer to the uniqueness of words to each topic. Topics falling in the upper right corner have the highest "quality" maximizing both exclusivity and semantic coherence.

A.2 Covariates, Sample & Results

The following section includes results that were discussed but not presented in the body of the paper. Topics presented in Table A.3 are produced with the exact same model as Table 1, the only difference is the second column where highest frequency word summaries for topics are presented in lieu of FREX word topic summaries.

Table A.1: Structural Topic Model Covariate Summary

Factor Level	Gender	Party	Electoral Experience
0	Male	Republican	Experienced Candidate or Incumbent
1	Male	Republican	Political Amateur
2	Female	Republican	Experienced Candidate or Incumbent
3	Female	Republican	Political Amateur
4	Male	Democrat	Experienced Candidate or Incumbent
5	Male	Democrat	Political Amateur
6	Female	Democrat	Experienced Candidate or Incumbent
7	Female	Democrat	Political Amateur

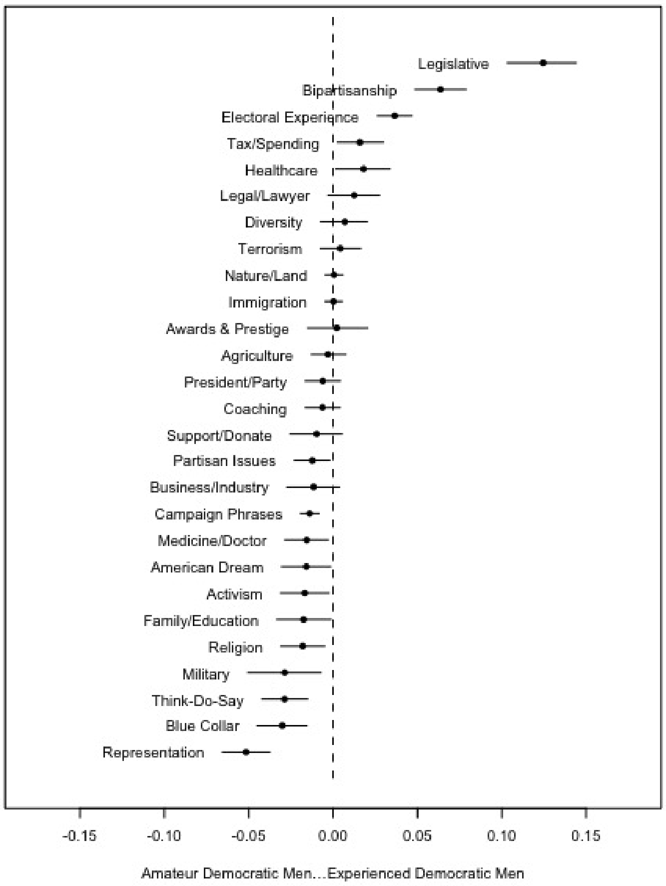
Table A.2: Percent of Candidates Across Each Race Type, By Gender

Gender	Republican Incumbent	Democrat Incumbent	Open Seat
Male	47.58% 275	26.99% 156	25.43% 147
Female	44.14% 162	25.89% 95	29.97% 110

Table A.3: Topics as a Function of Gender, Party, and Political Experience using Highest Frequency Words

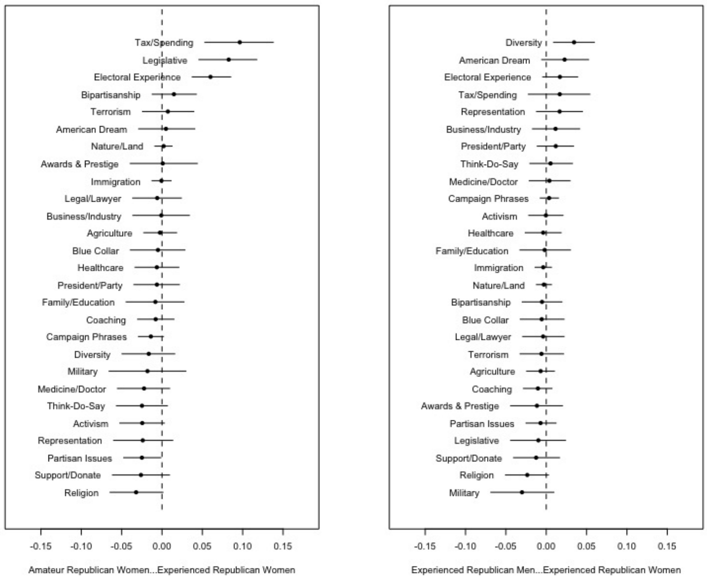
Description	Stems	%
Family / Education	school, work, year, colleg, famili, graduat, univers, high, rais, children	0.08
Awards / Prestige	univers, serv, member, board, school, associ, award, degre, presid, state	0.07
Representation	peopl, work, need, repres, believ, communiti, govern, district, want	0.07
American Dream	work, famili, busi, american, hard, small, job, dream, help, valu, start, communiti, opportun, grow, compani, generat, live, parent, creat, rais	0.06
Business / Industry	busi, develop, manag, experi, communiti, local, year, career, industri, econom, includ, leadership, compani, servic, project, director, public	0.05
Legislative	committe, hous, serv, legisl, member, congress, repres, also, congressman	0.05
Military	serv, servic, militari, armi, forc, unit, air, veteran, offic, state	0.05
Tax / Spending	tax, govern, conserv, busi, state, small, spend, budget, fight, right, nation	0.05
Think-Say-Do	can, get, like, time, just, make, peopl, know, one, live, need, thing, want, life, back, take, politician, come, even, say	0.05
Bipartisanship	job, work, help, congress, creat, veteran, improv, new, economi, make, secur, ensur, get, local, support, communiti, feder	0.04
Elect. Experience	district, elect, repres, serv, counti, citi, congression, state, year, member	0.04
Healthcare	health, care, educ, afford, healthcar, protect, access, right, qualiti, system	0.04
Activism	right, progress, fight, polit, corpor, chang, money, work, campaign, peopl, interest, power, worker, democrat, climat	0.03
Blue Collar	work, year, compani, manag, busi, engin, program, start, industri, time, experi, high, move, financi, mani, sever	0.03
Campaign Phrases	congress, run, district, time, america, home, now, meet, futur, congression	0.03
Diversity	communiti, educ, work, women, school, organ, advoc, program, student, public, help, children, support, first, state, democrat, polici, campaign, immigr	0.03
Medicine / Doctor	medic, health, care, doctor, research, univers, hospit, medicin, cancer, nurs, practic, patient	0.03
Legal / Lawyer	attorney, crimin, state, justic, offic, polic, counti, legal, crime, practic, serv	0.03
President / Party	republican, democrat, parti, polit, trump, presid, candid, campaign, american	0.03
Religion	year, life, state, serv, church, children, america, faith, unit, also, god	0.03
Support / Donate	vote, read, campaign, district, elect, donat, support, candid, pleas, repres	0.03
Agriculture	farm, counti, agricultur, volunt, farmer, cross, local, western, river, state, march, central, organ, valley, island, north, rural, involv, blue, fish	0.02
Coaching	play, team, coach, high, scout, sport, success, boy, tough, help	0.02
Terrorism	secur, nation, american, america, defens, presid, foreign, polici, war, world	0.02
Nature / Land	servic, face, resourc, water, open, parti, land, third, natur, generat, sourc, fourth, valley, light, enjoy, protect, bold, behalf, rang	0.01
Immigration	issu, paid, support, import, border, post, name, point, focus, top, congress, status, activ, divid	0.01
Partisan Issues	state, can, america, year, peopl, school, gun, unit, program, cost, need, use, provid, social, million, educ, must, tax, american, make	0.01

Figure A.5: Topic Prevalence by Partisanship Conditioned on Electoral Experience for Male Candidates



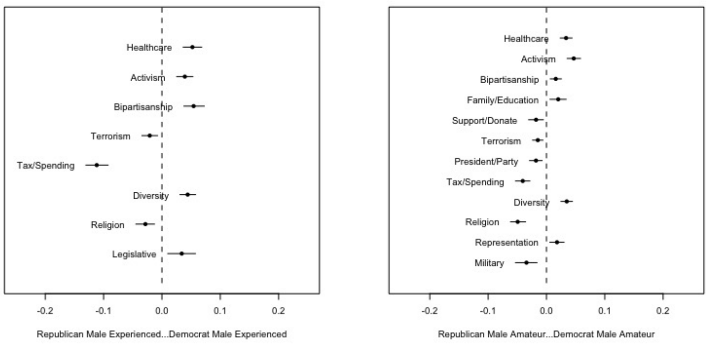
Note: The panel compares amateur Democratic men to experienced Democratic men. Null hypothesis is no difference in topic prevalence between candidate comparison groups. Point estimates are generated with 90% confidence intervals. Topics where the difference in prevalence was statistically indistinguishable from zero were omitted for clearer interpretation.

Figure A.6: Topic Prevalence for Experienced Republican Women vs. Comparison Groups



Note: The left panel compares amateur Republican women to experienced Republican women. The right panel compares experienced Republican men to experienced Republican women. The null hypothesis is no difference in topic prevalence between amateur and experienced candidates. Point estimates are generated with 90% confidence intervals.

Figure A.7: Topic Prevalence by Partisanship Conditioned on Electoral Experience for Male Candidates



Note: The left panel compares experienced male Democrats to experienced male Republicans. The right panel compares amateur male Democrats to amateur male Republicans. Null hypothesis is no difference in topic prevalence between candidate comparison groups. Point estimates are generated with 90% confidence intervals. Topics where the difference in prevalence was statistically indistinguishable from zero were omitted for clearer interpretation.