

ABSTRACT

Title of dissertation: ANALYZING COMMUNICATIVE CHOICES
TO UNDERSTAND THEIR MOTIVATIONS,
CONTEXT-BASED VARIATION, AND
SOCIAL CONSEQUENCES

Pranav Goel
Doctor of Philosophy, 2023

Dissertation directed by: Professor Philip Resnik
Department of Linguistics
Institute for Advanced Computer Studies

In many settings, communicating in a language requires making choices among different possibilities — the issues to focus on, the aspects to highlight within any issue, the narratives to include, and more. These choices, deliberate or not, are socially structured. The ever-increasing availability of unstructured large-scale textual data, in part due to the bulk of communication and information dissemination happening in online or digital spaces, makes natural language processing (NLP) techniques a natural fit for helping understand socially-situated *communicative choices* using that textual data. Within NLP methods, *unsupervised* NLP methods are often needed since digital large-scale textual data in the wild is often available without accompanying labels, and any existing labels or categorization might not be appropriate for answering specific research questions.

This dissertation seeks to address the following question: how can we use unsupervised NLP methods to study texts authored by specific people or institutions

in order to effectively explicate the communicative choices being made, as well as to investigate their potential motivations, context-based variation, and consequences?

Our **first** set of contributions centers on *methodological* innovation. We focus on topic modeling: a class of generally unsupervised NLP methods that can automatically discover authors' communicative choices in the form of topics or categorical themes present in a collection of documents. We introduce a new neural topic model (NTM) that effectively incorporates contextualizing sequential knowledge. Next, we find critical *gaps* in the near-universal automated evaluation paradigm that compares different models in the topic modeling methods literature, which calls into question much of the recent work in NTM development claiming “state-of-the-art” and emphasizes the importance of validating the outputs of unsupervised NLP methods.

In order to use unsupervised NLP methods to investigate potential motivations, context-based variation, and consequences of communicative choices, we link textual data with information about the authors, social contexts, and media involved in their production — these connected information sources help us conduct *empirical* research in social sciences.

In our **second** set of contributions, we analyze a previously unexplored connection between a politician's donors and their communicative choices in their floor speeches to show how donations influence issue-attention in US Congress, enabling a new look at money in politics and providing an example of studying *motivations* behind communicative choices.

Our **third** set of contributions uses text-based ideal points to better understand the role of institutional constraints and audience considerations in the *varying* expression and ideological positioning of politicians. The application of this tool for expanding knowledge of legislative politics is enabled by comprehensive annota-

tions for modeling outputs provided by domain experts in order to establish the tool's validity and reliability.

In our **fourth** set of contributions, we demonstrate the potential of both unsupervised NLP techniques and social network data and methods in better understanding the downstream *consequences* of communicative choices. We focus on misinformation narratives in mainstream media, viewing and highlighting misinformation as something that goes beyond just false claims published by certain bad actors or stories published by certain 'fake news' outlets. Our findings suggest a strategic repurposing of mainstream news by conveyors of misinformation as a way to enhance the reach and persuasiveness of misleading narratives.

ANALYZING COMMUNICATIVE CHOICES TO UNDERSTAND
THEIR MOTIVATIONS, CONTEXT-BASED VARIATION, AND
SOCIAL CONSEQUENCES

by

Pranav Goel

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:
Professor Philip Resnik, Chair/Advisor
Professor Kris Miler, Dean's Representative
Professor Jordan Boyd-Graber
Professor John Dickerson
Professor Naeemul Hassan
Professor David Lazer, External Member

© Copyright by
Pranav Goel
2023

Acknowledgments

Writing this, more than anything, makes me realize that I am coming to the end of this unique, challenging, and beautiful PhD journey. And it is clear to me that I would not have gotten here without the explicit support, mentorship, and camaraderie of so many wonderful people. Elucidating the enormous role they all played seems impossible to capture through words alone (ironic, given that text data and words drive much of my research). But I have to try. I owe it to all the people that made this particular journey a reality, and I owe it to the one reading this: ultimately, the fact that nothing is possible without the people in our lives and our connections with them is as important a takeaway as any in this dissertation.

I *have* to start with my wonderful advisor, Prof. Philip Resnik. Throughout the enormous variety of experiences and all the ups and downs in my PhD journey, Philip has always been that reason to feel happy, to feel incredibly lucky — a consistent bright spot no matter how dark the sky got. His kindness, energy, and immense support have all undoubtedly constituted the biggest positive experience of my PhD. Having a great advisor who is an awesome human being and having an extremely healthy relationship with them is not a given by any means; in fact, it is very clear to me that who the advisor is and the relationship with them largely determines the PhD experience, and is almost purely based in luck. I remember the agony and stress of not having a formal advisor during my first year (Philip was on a sabbatical). All of it was more than worth it to ultimately have Philip as my advisor. I will always fondly remember the moment after the end of my third semester when Philip formally offered to be my advisor. I wish I can capture some of what I have learned from Philip in my research, as a teacher (superb experiences with all the courses and seminars I have taken with him), and in being a mentor

myself to other scholars and people in my career. I wish I would be a mentor where people's energy levels and enthusiasm for both research and life in general *soared* after an hour of meeting and discussion with me — that this has been and still is my regular experience with Philip is absolutely remarkable.

So many other professors have inspired and supported me in myriad ways during my PhD. Prof. Jordan Boyd-Graber solidified my research interest in natural language processing, offered immense support during my transition period in my first year, and continually offered invaluable advice and help throughout my PhD. As my research interests shifted to computational political and social science, one would be hard-pressed to find better mentors and collaborators than Prof. Kris Miler and Prof. David Lazer. I learned a ton collaborating with all three different professors working in different disciplines. Taking classes taught by Prof. John Dickerson, Prof. David Jacobs, and Prof. Abhinav Shrivastava also helped expand my thought process during my early PhD years. As my research interests continue to evolve to include digital journalism information sciences, I have been considerably helped by the advice and perspective offered by Prof. Naeemul Hassan and Prof. Doug Oard during the final stretch of my PhD.

It is not a hyperbolic claim that friendships and community are vital to healthy functioning; without friends, absolutely nothing would have been possible for me. I have been, again, incredibly lucky to find wonderful friends in College Park. Sweta Agrawal has continually inspired me while being a constant pillar of strength and support — any amount of gratitude I could attempt to express would fall well short for her. From friends to roommates, it has been an honor to share the PhD experience with her from start to finish. Shlok Kumar Mishra and Susmija Jabbireddy have shown me, time and again, the power of people who care for you and always show up for you. Alexander Hoyle has both been my closest research colleague — a

cursory glance at my CV would confirm that — and been my cherished friend and comrade, almost like a big brother to me. Eleftheria Briakou has shown me what it means to be a great scientist while offering me stark moments of relatability and grounding me with her humor. The fact that my conversations with these people can move seamlessly between research and literally any other topic has been a source of both ease and joy. Plenty of others have kept me grounded and shared in the best and worst of my times: Sai Kanth Dacha, Sneha Gathani, Samyadeep Basu, Aman Geetey, Avira Rajesh, Rosie Brown, Erin Green, Nicholas Chu, Anshul Shah, Yogesh Balaji, Goonwanth Reddy Namireddy, Kamal Gupta, Ketul Shah, and Saumya Rawat. I am forever thankful to all these people.

My PhD experience has been markedly shaped by being part of the CLIP lab at UMD — it is hard to overstate the importance of sharing lab space, attending talks and colloquiums, and iterating on presentations and papers with people in the lab. The lab brings together a plethora of perspectives and a unique support system. Shi Feng was my research mentor during the entirety of my first year when I was without a formal faculty advisor and I will always look up to him. Sudha Rao, Pedro Rodriguez, and Joe Barrow further demonstrated how immensely helpful it is to have senior PhD colleagues as awesome mentors. I will seriously miss my colleagues and friends in the lab who have both uplifted and grounded me: Rupak Sarkar, Navita Goyal, Neha Srikanth, Maharshi Gor, Yang (Trista) Cao, Yoo Yeon Sung, Chenglei Si, Sathvik Nair, Shramay Palta, and Connor Baumler. I also could not have done without all the invaluable help with statistics always readily and happily offered by Peter Rankel.

Other colleagues and research collaborators have significantly added to my educational experience, taught me how to be a better researcher, and set an example as my mentors. I am so happy I got to collaborate, and will hopefully continue

to collaborate, with SoRelle Wyckoff Gaynor at UMD and now at the College of Holy Cross; Jon Green, Jason Radford, Sarah Shugars, Briony Swire-Thompson, and Stefan McCabe via the Lazer Lab at Northeastern University; and Nebojsa Jojic, Nikolay Malkin, and Sameera Lanka via Microsoft Research.

I have also found a powerful community brought together by the fight for better working conditions for graduate workers on the UMD campus and democratizing our academic workplace — my friends and comrades at the Fearless Student Employees of UMD, I will forever cherish the steps we have taken towards our goals, and I am convinced that even bigger wins are on the horizon. The fire and comradeship shown by Sam DiBella, Nate Beard, Jan-Michael Archer, and Erin Green, among others, will forever stay with me as an example of how to care for all our fellow academic workers through actions that refuse to settle for the hand we have been dealt, and to continually organize for a better future in the now.

I also thank the staff and admin at the Department of Computer Science and UMIACS for being so responsive to all my queries and concerns and continually providing a variety of vital support, from computing resources to financial and travel management. Special shoutout to the one-person powerhouse that is Tom Hurst for being the most reliable and kind go-to person; the entire graduate student body in CS at UMD falls apart without what he does so amazingly well on a daily basis. I am thankful for all the funding that carried me through my PhD: the teaching assistantships by the CS department, research grant money from the National Science Foundation (NSF) under grants 2008761 and 2031736, and additional funding support by Amazon AWS and via my summer internships at Microsoft Research.

A range of magnificent mentors and friends through my undergraduate years gave me the encouragement and faith needed to even begin my PhD journey:

incredible senior scholars in Aditya Joshi and Yoichi Matsuyama, who were my first and closest research mentors and set an example I will continually try to emulate; Prof. Anil Kumar Singh, my very first faculty mentor who provided me concrete support as I took my first steps in learning and doing research in NLP; Prof. Pushpak Bhattacharyya, who gave me an incredible research internship opportunity as a sophomore just starting to get genuinely interested in NLP research; Prof. Justine Cassell, who took a chance on a junior undergraduate in India and brought him on to work at her amazing lab at Carnegie Mellon in the US for a summer research internship and introduced me to other wonderful colleagues in Vivian Tsai and Florian Pecune, influential mentors Michael Madaio and Samantha Finkelstein, and lab manager extraordinaire Lauren Simmons. The invaluable base of research experience as an undergraduate and the arduous, enriching process of graduate school applications would not have been possible without the fun collaborations, discussions, and joint effort with Devang Kulshreshtha, Prayas Jain, Mayank Rathee, Ayush Shrivastava, Anant Dadu, and Prateek Thakur.

My existing and new close friendships beyond College Park have kept me going through it all, and I am extremely fortunate to have these phenomenal people as part of my hopefully perennial support network: Shubham Chaudhary, Raghav Dhawan, Mahika Phutane, and Grusha Agarwal.

Nothing, from the journey leading to the start of my PhD, my time and any successes during my PhD, and what happens beyond, would be possible without my parents, my grandparents, my sister, and the wondrous extended family I had the pleasure to grow up and share my home with. I thank them for their spectacular patience and for always putting my happiness first.

Table of Contents

Acknowledgements	ii
Table of Contents	vii
List of Tables	xiii
List of Figures	xxxii
1 Introduction	1
1.1 Contributions	4
1.2 Roadmap for this Dissertation	6
2 Extracting Communicative Choices from Textual Content: Enhancing and Comparing Topic Models	8
2.1 Incorporating Contextual Knowledge in Neural Topic Models	10
2.1.1 Method	13
2.1.2 Coherence-based Automated Evaluation	18
2.1.3 Experimental Details and Results	21
2.1.4 Topic Alignment based Comparison and Analysis	26
2.1.5 Main Takeaway	29

2.1.6	Related Work	31
2.2	Automated Coherence-based Evaluation of Topic Models is Broken	33
2.2.1	Meta-analysis of Neural Topic Modeling	38
2.2.2	Human Measures of Topic Coherence	41
2.2.3	Experimental Detail	42
	Datasets and Preprocessing	42
	Models	43
2.2.4	Process for Conducting Human Evaluation	45
2.2.5	Results: Human Judgments Differ from Automated Metrics	47
2.2.6	Main Implications	54
3	Motivations behind Communicative Choices: Donor Activity is Associated with US Legislators’ Attention to Political Issues	57
3.1	Introduction and Background	59
3.2	Dataset Overview	65
3.3	Methodological Framework	69
3.4	Experimental Detail	73
3.4.1	Data Processing	73
3.4.2	Topic Modeling	75
3.4.3	Regularized Multinomial Logistic Regression Framework and Training Detail	76
3.4.4	Measuring the Association	78
3.5	Results	80
3.6	Expert Assessment of PAC-Issue Associations Learned by Our Model	84
3.7	Procedure for Surfacing Potential Temporal Connections	86
3.8	Discussion and Conclusion	89

4	Context-based Variation in Communicative Choices: US Legislators' Ideological Expression Across Audiences	95
4.1	Introduction	98
4.2	Background	100
4.3	Theoretical Motivation: The Case for Multiple Measures of Ideology	106
4.4	Data	108
4.5	Method	110
4.6	Results and Analyses	117
4.7	Discussion and Conclusion	131
4.8	Potential Limitations and Caveats	135
5	Consequences of Communicative Choices: Mainstream News Articles Co-Shared with Fake News Buttress Misinformation Narratives	137
5.1	Introduction	138
5.2	Background and Definitions	142
5.3	Data and Methods	148
5.4	Hypothesis Testing Procedure and Results	156
5.5	Case Study	160
5.6	Discussion and Related Work	175
5.7	Limitations and Future Work	178
5.8	Potential Broader Impact	182
6	Conclusion and Future Research Agenda	184
6.1	Conclusion and Summary of Contributions	184
6.2	Future Research Directions	189
6.2.1	Systematic <i>Exclusion</i> of Certain Communicative Choices	189

6.2.2	Beyond Individual Choices: Using Similarities in Communicative Choices to Model Agreement in Decision-Making for Pairs of Actors	194
6.2.3	Studying Variation in Ideological Expression across Political Issues	201
6.2.4	Estimating Causal Association Between Donation Activity and Legislators' Communicative Choices	205
6.2.5	Investigating Possible Contrasting Choices in Opinion Pieces (Op-Eds) and News Published by the Same Popular Mainstream Media Outlets	206
6.2.6	Does Misinformation Benefit from a Partnership with Expression of Targeted Hatred?	209
A	Extracting Communicative Choices from Textual Content: Enhancing and Comparing Topic Models	211
A.1	Dev Set Results	211
A.2	Extrinsic Classification Results	212
A.3	Using BAT with DVAE	213
A.3.1	Topic Diversity	214
A.4	Experimental Procedures	215
A.4.1	Hyperparameter Tuning and Optimal Values	215
A.4.2	Computing Infrastructure and Runtime	216
A.5	Changes to W-LDA	219
A.6	Impact of BAT on Individual Topics: Aligned Topic Pair Examples	220
A.7	List of Neural Topic Modeling Works used in our Meta-Analysis	222
A.8	Preprocessing Details	226

A.9	Training Details	227
A.10	Instructions for Crowdworkers	230
A.11	Regression Results	231
A.12	Filtering on Term Familiarity	232
A.13	Five-point Ratings Scale	234
A.14	Potential Negative Impact	235
B	Motivations behind Communicative Choices: Donor Activity is Associated with US Legislators’ Attention to Political Issues	236
B.1	Additional Experimental Detail	236
B.2	Additional Results	237
B.3	Comparing <i>PAC</i> and <i>Committee</i> Legislator Attributes	243
B.4	PAC-Issue Association Judgment Task Details for Subject-Matter Experts	245
B.5	More Examples and Results of Discovered Temporal Connections . .	249
C	Context-based Variation in Communicative Choices: US Legislators’ Ideo- logical Expression Across Audiences	251
C.1	Preprocessing Details	251
C.2	Stability of Ideal Point Estimates	252
C.3	Details of Expert Assessment Tasks to Validate Automated Outputs .	253
C.3.1	Initial Topic Assessment (Pre-TBIP)	254
C.3.2	Post-TBIP Output Assessment — Ideological Topics	255
C.3.3	Summary of results for floor speeches	256
C.3.4	Summary of results for tweets	257
C.4	Extended Results	258

D	Consequences of Communicative Choices: Mainstream News Articles	
	Co-Shared with Fake News Buttress Misinformation Narratives	261
D.1	Distribution of Article Counts per Outlet in Co-shared & Control Group	261
D.2	Partisanship Explanation Analysis: Procedure and Results	261
D.3	Case Study Details: All Narrative Labels after Keyword-based Filtering	266
D.4	Case Study Details: Full Unique Tweet Text Set by Vaccine Misinformation Sharers	266
D.5	Case Study Details: Measuring the Impact of Type-1 Mainstream Articles in Terms of Uptake amongst Vaccine Misinformation Sharers	271
D.6	Direct Quotation Analysis: Procedure and Results	272
D.7	Automatically Extracted Narratives from False Fact-checked Claims .	274
	Bibliography	296

List of Tables

1.1	The types of communicative choices explored in this dissertation. . . .	3
1.2	Summary of the various broad types of contributions made in this dissertation. <i>Validation of existing methods</i> includes critical examinations of existing methods or evaluation paradigms in methods research as well as assessment of outputs of automated unsupervised methods in order to understand their reliability. <i>New data</i> also includes new data generated from outputs of our research, which provides new labels and enables new studies stemming from our findings and their context (for example, grouping mainstream ‘real’ news articles into categories that signify if they share synergy with ‘fake’ news).	4
2.1	Corpus statistics, which vary in total number of documents (D), vocabulary size (V), and average document length (N_d).	21
2.2	The NPMI for our baselines (Section 2.1.3) compared with BAT (explained in Section 2.1.1) using SCHOLAR as our base neural architecture. We achieve better NPMI than all baselines across three datasets and $K = 50, K = 200$ topics. We use 5 random restarts and report the standard deviation.	24

2.3	Mean NPMI (s.d.) across 5 runs for W-LDA Nan et al. (2019) and W-LDA + BAT for $K = 50$, showing improvement on two of three datasets. This demonstrates that our method is <i>modular</i> and can be used with base neural topic models that vary significantly in architecture.	26
2.4	External NPMI (s.d.) for the base SCHOLAR and SCHOLAR + BAT. Models selected according to performance on the development set using internal NPMI.	26
2.5	Selected examples of SCHOLAR + BAT improving on topics from SCHOLAR. We observe that the improved 20ng topic is more cleanly focused on the NHL (removing <i>european</i> , adding the Toronto Maple Leafs, evoking the Stanley Cup rather than the more generic <i>ice</i>); the improved wiki topic about typhoons is more clearly concentrated on meteorological terms, rather than interspersing specific locations of typhoons (<i>luzon, guam</i>); and the improved IMDb topic more cleanly reflects what we would characterize as “video adaptations” by bringing in terms about that subject (<i>book, books, novels, read</i>) in place of predominant words relating to particular adaptations. Randomly selected examples can be found in Appendix A.6.	30

2.6	The first three columns are the highest-NPMI topics for a classical topic model (LDA estimated via Gibbs sampling using Mallet, McCallum, 2002; Griffiths and Steyvers, 2004). The next three are counterparts from a neural model (our D-VAE reimplementation, Burkhardt and Kramer, 2019). Models are trained on Wikitext (Merity et al., 2017) with fifty topics, and NPMI is estimated over the top five words in each topic using a 4.6M-document reference Wikipedia corpus. The mean top-five NPMI over all topics is 0.156 for the classical and 0.256 for the neural model.	36
2.7	Meta-analysis of forty neural topic modeling papers (denominator may change, as not all conditions are applicable). No recent neural topic modeling papers use human evaluations of coherence, and the metrics and models are difficult to replicate.	38
2.8	Corpus statistics. Datasets vary in domain, average document length, and total number of documents. WIKI is from Merity et al. (2017) and NYT is from Sandhaus (2008).	42
2.9	Spearman correlation coefficients between mean human scores and automated metrics. <u>Underlined</u> values have overlapping bootstrapped 95% confidence intervals with that of the largest value in each row. “Concatenated” refers to correlations computed on a concatenation of values for the NYT and WIKI items. “Val” is a small held-out set of 15% of the training corpus. Using the more data-appropriate logistic and ordered probit regressions for word intrusion and ratings data leads to different conclusions about relative metric strength (appendix table A.16). CIs are estimated using 1,000 samples.	49

2.10	False discovery rate (1 – precision, lower is better) and false omission rate of significant model differences when using automated metrics; automated metrics often overstate meaningful model differences. Bolded values are those with the lowest geometric mean of FDR and FOR. We sample two independent sets of 50 topics along with their human scores and automated metrics; these sets act as the outputs of two “models”. We then compute significance tests between sets (per Fig. 2.8) on both the automated scores and human scores. A false positive occurs when one set has significantly larger automated scores despite no meaningful difference in actual human scores. Estimates are over 1,000 samples.	52
2.11	Topics with the largest human–NPMI discrepancies; top half are topics where NPMI is high and human preferences are low, bottom half is the reverse. NPMI favors esoteric and corpus-specific topics. NPMI is calculated with a 10-token sliding window over the in-domain reference corpus, Rat. is the average 3-point rating for a topic, and Int. refers to the percentage of annotators who identify the intruder word.	52

3.1	The number of speeches and legislators in our processed dataset, along with the sizes of each of the legislator attributes used as possible explanatory variables in our study (<i>Random-PAC</i> is the same as <i>PAC</i>). The attributes (or legislator representations) are described in Section 3.3. These details are applicable to our model trained on the entire dataset; we provide these attribute sizes for models trained separately on data from each congressional cycle in Table B.2. We note that the six other possible values for <i>State</i> are: American Samoa, District of Columbia, Guam, Northern Mariana Islands, Puerto Rico, and Virgin Islands.	73
4.1	OLS regression results for Democratic party ideal points across venues, 115th – 116th Congress. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$; MC = Member of Congress.	120
4.2	OLS regression results for Republican party ideal points across venues, 115th – 116th Congress. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$; MC = Member of Congress.	121
4.3	OLS regression results for change in ranked ideal point positions across venues, 115th – 116th Congress. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$; MC = Member of Congress.	124
5.1	Misinformation narrative libraries.	154
5.2	Most frequent high-dimensional narrative strings or labels (and sentences containing those labels) in misinformation content, that are present in co-shared articles but not present in control articles published by mainstream ‘real’ news outlets.	155

5.3	Results of significance testing showing that across various categories of reliable mainstream news outlets and sets of misinformation narratives, co-shared articles have a significantly higher presence of misinformation narratives compared with the control group.	157
5.4	Logistic regression results for $y \sim X_{\text{misinfo_narrative}}$, where $y = 1$ when the mainstream article is co-shared, and $y = 0$ when the mainstream article is control. $X_{\text{misinfo_narrative}}$ is the count for the number of narratives in the mainstream article that are also present in the misinformation narrative set. The results show that an article having a higher number of misinformation narratives is significantly and positively associated with co-sharing likelihood. The estimate can be roughly interpreted as: the co-shared group has 1.25 times the odds of the control group for containing a higher number of misinformation narratives. *** indicates significance at $p < 0.001$; and $1.25 \approx \exp^{0.2241}$	159
5.5	Part one of the full set of selected narrative labels capturing the type of narrative we examine in our case study.	162
5.6	Part two of the full set of selected narrative labels capturing the type of narrative we examine in our case study.	163
5.7	Final part (part three) of the full set of selected narrative labels capturing the type of narrative we examine in our case study.	164
5.8	Part one of all the articles (qualitatively categorized as type 1) published by liberal outlets that are co-shared with fake news and contain selected narratives for our case study examining a particular type of narrative of the form ‘vaccinated people spread (virus, disease, variant, etc.).’	166

5.9	Part two of all the articles (qualitatively categorized as type 2) published by liberal outlets that are co-shared with fake news and contain selected narratives for our case study examining a particular type of narrative of the form ‘vaccinated people spread (virus, disease, variant, etc.).’	167
5.10	The final part (part three) of all the articles (qualitatively categorized as type 3) published by liberal outlets that are co-shared with fake news and contain selected narratives for our case study examining a particular type of narrative of the form ‘vaccinated people spread (virus, disease, variant, etc.).’	168
5.11	Vaccine-related fake news significantly co-shared with identified liberal mainstream articles in our case study.	171
5.12	Sample of tweets texts when vaccine-misinformation-conveyors on Twitter share particular mainstream articles identified in our case study.	173
6.1	Preliminary results for our basic mixed-effects model (Eq. (6.1)) using just three pairwise features or fixed effects (X_{ij}): same party, same state, and the cosine similarity between the TF-IDF vectors formed from the text of floor speeches for the pair of legislators. The dependent variable is the co-vote agreement rate for the 109 th Congress. The estimates indicate that even using very basic speech-based features can offer a meaningful association and explanation for the co-vote rate between two legislators. As expected, the more similar the speech vectors, the higher the rate of voting together.	198

6.2	Additional preliminary results for our mixed-effects model (Eq. (6.1)) using just four pairwise features or fixed effects (X_{ij}): same party, same state, the cosine similarity between the TF-IDF vectors formed from the floor speeches for the pair of legislators, and similarity between the estimated ideal point values for legislators derived from their floor speeches using the TBIP method. The dependent variable is the co-vote agreement rate for the 109 th Congress. The estimates indicate that the closeness of speech-based ideal point values for legislators has a significant impact on the co-vote agreement rate, more than sharing the same party or state and even more than our basic modeling of speech-based legislator similarity using TF-IDF vectors. This further shows the promise of using language-based similarity between decision-makers, and how more sophisticated language-based tools can add further value to co-decision modeling.	199
A.1	The development-set NPMI for our baselines (Section 2.1.3) compared with BAT (explained in Section 2.1.1) using SCHOLAR as our base neural architecture. We achieve better NPMI than all baselines across three datasets and $K = 50, K = 200$ topics. We use 5 random restarts report the standard deviation.	211
A.2	The mean development-set NPMI (std. dev.) across 5 runs for W-LDA and W-LDA + BAT for $K = 50$, showing improvement on all datasets. This demonstrates that our innovation is <i>modular</i> and can be used with base neural topic models that vary in architecture.	211
A.3	Random forest classification accuracy on 20ng and IMDB datasets, using topic estimates from SCHOLAR and SCHOLAR + BAT.	212

A.4	Mean development set NPMI (s.d.) across 5 runs for DVAE (Burkhardt and Kramer, 2019) and DVAE + BAT for $K = 50$	214
A.5	Hyperparameter ranges and optimal values (as determined by development set NPMI) for SCHOLAR and SCHOLAR+ BAT, on the 20NG dataset. lr is the learning rate, α is the hyperparameter for the logistic normal prior, λ is the weight on the teacher model logits from Eq. (2.4), and T is the softmax temperature from Eq. (2.4). Other hyperparamter values (which can be accessed in our code base) which were kept at their default values are not reported here. Values marked with the * are also kept at their default values per the base SCHOLAR model (https://github.com/dallascard/scholar). All different sweeps in the grid search were run for 500 epochs with a batch size = 200	216
A.6	Hyperparameter ranges and optimal values (as determined by development set NPMI) for SCHOLAR and SCHOLAR+ BAT, on the Wiki dataset. lr is the learning rate, α is the hyperparameter for the logistic normal prior, anneal controls the annealing (as explained in Appendix B in Card et al. (2018)), λ is the weight on the teacher model logits from Eq. (2.4), T is the softmax temperature from Eq. (2.4), and clipping controls how much of the logit distribution to clip (Section 2.1.1). Other hyperparamter values (which can be accessed in our code base) which were kept at their default values are not reported here. All different sweeps in the grid search were run for 500 epochs with a batch size = 500	217

A.7	Hyperparameter ranges and optimal values (as determined by development set NPMI) for SCHOLAR and SCHOLAR+ BAT, on the IMDb dataset. lr is the learning rate, α is the hyperparameter for the logistic normal prior, anneal controls the annealing (as explained in Appendix B in Card et al. (2018)), λ is the weight on the teacher model logits from Eq. (2.4), T is the softmax temperature from Eq. (2.4), and clipping controls how much of the logit distribution to clip (Section 2.1.1). Other hyperparameter values (which can be accessed in our code base) which were kept at their default values are not reported here. Values marked with the * are also kept at their default values per the base SCHOLAR model (https://github.com/dallascard/scholar). All different sweeps in the grid search were run for 500 epochs with a batch size = 200.	217
-----	--	-----

A.8	Hyperparameter ranges and optimal values (as determined by development set NPMI) for W-LDA and W-LDA+ BAT, on all three datasets. lr is the learning rate, α is the hyperparameter for the dirichlet prior, λ is the weight on the teacher model logits from Eq. (2.4), T is the softmax temperature from Eq. (2.4), and clipping controls how much of the logit distribution to clip (Section 2.1.1). Other hyperparameter values (which can be accessed in our codebase) which were kept at their default values in the original baseline code are not reported here (also see Nan et al. (2019) and https://github.com/aws-labs/w-lda/). Values marked with the * are also kept at their default values. All different sweeps in the grid search were run for 500 epochs and noise parameter = 0.5 (see Nan et al. (2019)). For 20NG and IMDB, we used batch size = 200 , and for Wiki, we used batch size = 360 .	218
A.9	For DVAE, we tried four values for the Dirichlet Prior (as per the values tried by the authors in Burkhardt and Kramer (2019)) - $\{0.01, 0.1, 0.2, 0.6\}$ and report the optimal values corresponding to the dev set results (Table 2.2) and test set results (Table A.1) in this table. Within the model variations available in the codebase for DVAE (https://github.com/sophieburkhardt/dirichlet-vae-topic-models) we use the Dirichlet VAE based on RSVI which is shown to give the highest NPMI scores in Burkhardt and Kramer (2019).	219
A.10	Fifteen aligned topic pairs from the 20NG dataset.	222
A.11	Fifteen aligned topic pairs from the Wiki dataset.	223
A.12	Fifteen aligned topic pairs from the IMDB dataset.	224
A.13	Papers used in meta-analysis, Section 2.2.1	225

A.14	Hyperparameter settings for G-LDA, D-VAE, and ETM. *: Best setting for WIKI, †: best setting for NYT; based on NPMI estimated with a 10-token sliding window over the reference corpus.	229
A.15	Runtimes for the three topic models on each of the two datasets. G-LDA requires CPUs only while the neural models use a single GPU. Compute resources detailed at the end of Appendix A.9.	230
A.16	Logistic (intrusion) and ordinal probit (ratings) regression coefficients of automated metrics on human annotations. <u>Underlined</u> values have overlapping 95% confidence intervals with that of the largest value in each row.	232
A.17	Tables 2.9, 2.10, and A.16 after removing respondents who report a lack of familiarity with topic words.	233
A.18	Spearman correlation coefficients between mean human scores for a five-point ratings scale (rather than three), compared to table 2.9. <u>Underlined</u> values have overlapping 95% confidence intervals with that of the largest value in each row.	234
B.1	Values for optimal hyperparameter settings for the multinomial logistic regression model. α = learning rate; λ = L_2 regularization constant or the weight decay; η = binary indicator of whether to use early stopping (1) or not (0). The best settings found for PAC were also used for Random-PAC.	237

B.2	The sizes or dimensionality of each of our legislator attributes or representations for when the models are trained separately for each Congressional cycle. The first value in each dimensionality is the number of speakers selected in that cycle, and the second value is the number of features (for the speaker representation or model indicated in the column title). Note that the y for each of the regression models (averaged topic distribution for speakers) is going to be (number of speakers, 30).	237
B.3	Statistical significance test results using Mann-Whitney U test, comparing 30-fold cross-validation results for <i>PAC</i> against other legislator attributes. $p < 0.05$ (boldfaced) indicates significantly better results for the PAC attribute against the other attribute for that congressional session.	240
B.4	Statistical significance test results using Mann-Whitney U test, comparing 30-fold cross-validation results for <i>Committee</i> against other legislator attributes. $p < 0.05$ (boldfaced) indicates significantly better results for the Committee attribute against the other attribute for that congressional session.	241
B.5	Note: Experts (or annotators) created labels for topic using not just the top terms but the top documents (floor speeches) for the topic given by the topic model. And contextualizing info such as sector, industry, category was provided when asking 1-3 associations ratings.	246

B.6	After selecting relevant PAC-issue pairs with potential temporal connection for the cycle: some examples of speeches made on the topic and cases of the relevant PAC making a donation much higher than their mean donation amount during that congressional cycle to the legislator giving that speech within a particular time window around the speech.	249
C.1	Initial topics used to initialize TBIP, rated for coherence by two subject-matter experts.	254
C.2	Initial topics used to initialize TBIP rated at least 2 on the 3-point Likert scale for coherence, then rated for expected ideological polarization by two subject-matter experts.	254
C.3	Rating if the provided topic label (created using initial topics) applies to the the binary ideological perspective (<i>a</i>) for that same topic post-TBIP.	255
C.4	Rating if the provided topic label (created using initial topics) applies to the the binary ideological perspective (<i>b</i>) for that same topic post-TBIP.	255
C.5	Rating if the two ideological perspectives on an issue derived using TBIP represent polarized or partisan framings of the issue.	255
C.6	Full Congressional OLS model, 115th – 116th Congress. Party (Republican) acts as an interaction variable for those relevant variables (any independent variable that applies to both Democratic and Republican members). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$; MC = Member of Congress.	259

C.7	Changes in Ranked Ideal Point Position Interactive OLS model, 115th – 116th Congress. <i>Negative sign indicates more conservative than vote, positive sign indicates more conservative than vote. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$; MC = Member of Congress.</i>	260
D.1	Results of significance testing showing that across various categories of reliable mainstream news outlets when using our graph-based method and entire method (Section 5.3) starting with the subset of all Twitter data where tweet authors are conservatives (registered Republicans) only.	263
D.2	Logistic regression results for $y \sim X_{\text{partisanship}} + X_{\text{misinfo_narrative}}$, where $y = 1$ when the mainstream article is co-shared, and $y = 0$ when the mainstream article is control. A higher partisanship score means more conservative. The results show that an article being more conservative (measured by relative popularity among Republicans and Democrats in terms of their sharing behavior on Twitter) and an article having a higher number of misinformation narratives are significantly and positively associated with co-sharing likelihood. *** indicates significance at $p < 0.001$	264

D.3	We run two logistic regression models, the full model where both factors are used as independent variables ($y \sim X_{\text{partisanship}} + X_{\text{misinfo_narrative}}$), and partisanship-only model ($y \sim X_{\text{partisanship}}$); where $y = 1$ when the mainstream article is co-shared, and $y = 0$ when the mainstream article is control. We compare the model fit using the information-theoretic concept of Akaike information criterion (AIC), implemented using the R library ‘AICcmodavg’ (Mazerolle, 2023). In the table, K is the number of parameters of the model, AICc is the information score of the model (lower indicates better fit), Delta_AICc is the difference in the AIC score between the best model and the model being compared, and AICcWt is the proportion of the total amount of predictive power provided by the full set of models contained in the model being assessed (see: https://www.scribbr.com/statistics/akaike-information-criterion/ for more details).	265
D.4	All automatically extracted narrative labels or strings in liberal co-shared articles containing the terms ‘vaccin’ as well as ‘spread’. . . .	267
D.5	Results of significance testing showing that across various categories of reliable mainstream news outlets after all sentences containing direct quotes (signified through quotation markers) were removed. Across various sets of misinformation narratives, co-shared articles have a significantly higher presence of misinformation narratives compared with the control group.	273
D.6	All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3). .	275

D.7	All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3);	
	<i>cont.</i>	276
D.8	All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3);	
	<i>cont.</i>	277
D.9	All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3);	
	<i>cont.</i>	278
D.10	All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3);	
	<i>cont.</i>	279
D.11	All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3);	
	<i>cont.</i>	280
D.12	All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3);	
	<i>cont.</i>	281
D.13	All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3);	
	<i>cont.</i>	282
D.14	All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3);	
	<i>cont.</i>	283

D.15 All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); <i>cont.</i>	284
D.16 All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); <i>cont.</i>	285
D.17 All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); <i>cont.</i>	286
D.18 All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); <i>cont.</i>	287
D.19 All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); <i>cont.</i>	288
D.20 All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); <i>cont.</i>	289
D.21 All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); <i>cont.</i>	290
D.22 All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); <i>cont.</i>	291

D.23 All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); <i>cont.</i>	292
D.24 All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); <i>cont.</i>	293
D.25 All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); <i>cont.</i>	294
D.26 All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); <i>cont.</i>	295

List of Figures

2.1	We compare the usage of BERT and LDA models to emphasize the uptake and impact of topic modeling methods in research conducted outside the field of computer science. This comparison is done using research works between 2019 and 2021 (using just one topic modeling method: LDA). The data on citations and research areas of the publications come directly from Semantic Scholar (https://www.semanticscholar.org/), and the % of publications extracted in each area that cite BERT (Devlin et al., 2019) or LDA (Blei et al., 2003) are shown.	9
2.2	Improving a base neural topic model with knowledge distillation. A document is mapped through both a standard BoW representation and a BERT-based Auto-encoder “Teacher” (BAT), yielding two distributions over words. These are used as the ground truth in the “student” topic model’s document reconstruction loss $\mathcal{L}_{KD}$ (backpropagated along the dotted line). Crucially, the BAT distribution assigns mass to unobserved but related terms (unbolded).	11
2.3	Jensen-Shannon divergence for aligned topic pairs in the SCHOLAR and SCHOLAR + BAT models.	28

2.4	Number of matched topic pairs where SCHOLAR + BAT improves coherence, compared with the number of matched pairs where the baseline improves coherence.	29
2.5	An informal illustration of the fact that the reported ranges of coherence (such as NPMI) results is well outside the range of variable comparison in Lau et al. (2014), which used non-neural topic modeling techniques (proposed before 2010) in order to validate automated coherence metrics with human judgments—we do not know if the same relationship that was found would hold in this new range of values (achieved using contemporary neural methods proposed after 2015).	35
2.6	The word intrusion task presented to crowdworkers (the ratings task is shown in Fig. 2.7).	46
2.7	Ratings task presented to crowdworkers.	46
2.8	While automated evaluations (here, NPMI) suggest a clear winner between models, human evaluation is more nuanced. Human judgments exhibit greater variability over a smaller range of values. Colored circles correspond to pairwise one-tailed significance tests between model scores at $\alpha = 0.05$; for example, the rightmost orange circle at the bottom right shows that human intrusion ratings for D-VAE are significantly higher than ETM for topics derived from Wikipedia.	48
2.9	Mean human evaluation on the ratings and word intrusion tasks, after filtering out respondents who reported a lack of familiarity with the topic words. When filtering, D-VAE scores improve, highlighting its tendency to produce esoteric topics.	53

3.1	Our new dataset’s schema. Our main data contribution relies on creating a mapping between the <code>bioguide_id</code> , which is the primary key or ID for legislators in congressional Record Data providing floor speeches transcripts as well as the repository providing the accompanying biographical and committee membership information), and the <code>opensecrets_id</code> , which is the primary key or ID used to refer to political candidates receiving donations in the OpenSecrets database providing campaign contributions transactions data. Note that the columns or information shown for each table is not exhaustive; this is just an overview of the structure and contents of our database. The <code>pac_id</code> is originally found in FEC data and used in the bulk data made available by OpenSecrets that we use. There are 5 possible values for the type of PAC: Business, Labor, Ideological, Other, and Unknown — we only consider Business and Labor PACs in our study. All the various ID mappings in our database are either one-to-one or one-to-many, represented using ‘1’ and ‘*’ on the mapping links in the schematic diagram.	66
3.2	Our model automatically quantifies disproportionate attention paid by legislators to the issues relevant to the PACs donating to them — this is illustrated here using a simulated legislator as an example, using the actual average topic distribution across all legislators (and their speeches in our data) and the legislator’s topic distribution per our model. With healthcare and education-related PACs donating to the legislator, the model trained to capture associations between donors and issue-attention expects a disproportionate focus on education and healthcare topics.	70

3.3	In our processed dataset covering 758 selected legislators giving floor speeches across 1995-2018, and 1002 PACs actively donating in the same period; most PACs tend to donate to less than a third of all the legislators they could donate to.	73
3.4	In our processed dataset used to train our machine learning framework, most PACs tend to donate across party lines instead of donating in an exclusive, partisan manner.	74
3.5	PACs are most associated with legislators' issue-attention in floor speeches compared with other legislator attributes, both in aggregate terms across the entire time period of our dataset as well as in individual congressional cycles. (a): shows the results in aggregate, with information about which individual PACs donated to legislators clearly explaining their issue-attention the most out of all the legislator attributes we consider; the cross-validation validation procedure was repeated 50 times, and the error bars (for results across the cross-validation procedure) are shown. (b): shows the results when repeating our topic modeling and machine learning procedure separately for each of the twelve congressional cycles in our data, and demonstrates the same pattern as the aggregate picture (with the cross-validation procedure repeated 30 times; the error bars are provided in Fig. B.3). (R) and (D) indicate which party was in the majority in the US House during that congressional cycle. Both of our findings are statistically significant (Appendix B.2).	81

3.6	Results of annotating PAC-issue associations, showing that the top PACs per issue as per our regression model’s weights are meaningful since the model’s top 10 PACs are rated consistently higher in terms of association with the issue as compared with a random selection of 10 PACs for the issues.	83
3.7	Algorithmic workflow of identifying relevant PAC-issue associations showing cases of significantly connected speech and donation events — cases where within a short time window (proximal), a donation made by a PAC interested in a particular issue and a speech made by the recipient on that issue co-occur.	86

3.8 An example of the kind of real-world relationships that the outputs of our modeling and analysis can capture automatically — for a PAC (*Career Education Colleges & Universities*) and an issue (*Education*) that our model (as well as experts) deem related to one another, we find that in a particular congressional cycle (2015-16), this PAC is significantly likely to donate a more-than-their-usual amount of money closer to dates on which speeches prioritizing the particular issue of education were made in the 2015-16 cycle. We found that such a significant connection is primarily driven by donations made to a particular legislator: *John Kline (R-MN)*. This legislator made a number of speeches relating to the issue of education on Feb 25, 2015 and Feb 26, 2015 – manually looking at the speeches, the legislator is arguing against federal funding for education and schooling and in favor of charter schools. The PAC made a donation of \$5000 to this legislator on Feb 27, 2015 – when the PAC’s average (std dev) donation amount to *any* legislator in that particular cycle was \$1862.16($\pm$ \$1288.27). These kinds of relationships can be uncovered by journalists, but it can be a time-intensive process. Our data plus modeling can find such relationships (by ranking them towards the top) without supervision, and therefore, our work can enable journalists to focus on cases deemed significant by our analyses as starting points for further investigation. Snapshots in the image come from <https://www.usatoday.com/story/news/politics/2013/07/23/for-profit-colleges-contributions-house-regulations/2579041/> and the [Wikipedia article](#) for Career Education Colleges and Universities. . . . 90

4.1	Illustration of “Liberal” and “Conservative” word usage on the issue of Abortion using our implementation of the TBIP method on our floor speeches data; generated using https://github.com/keyonvafa/tbip . The length of the bars indicates the strength (polarity) of the term v in topic k (η_{kv}).	113
4.2	Member distribution of the three ideal point measures, 115 th - 116 th Congress.	117
4.3	Categorical shift in member rankings from vote-based to text-based (floor speeches and tweets) ideal points, 115 th and 116 th Congress. . .	119

5.1	Visual illustration summarizing our approach. In part a) (top-left), we show our process of finding reliable articles disproportionately co-shared with fake news: we score pairs of fake and reliable news on the likelihood of co-occurrence of their shares (by the same set of Twitter users) when controlling for the individual popularity of the two articles. Our method uses a graph-based approach with edges between fake and reliable articles (the nodes) weighted by the number of users sharing both incident articles. The Washington Post article (only headline shown here) illustrates our approach: the article was disproportionately shared by anti-vaccine fake-news conveyors in our Twitter panel. This process helps us create two groups of mainstream news articles: disproportionately co-shared and control . Using a natural language processing method developed in prior work, we automatically extract potential narrative structures from all texts in our data: misinformation content as well as mainstream news articles, and narratives from misinformation content are considered sets of misinformation narrative libraries (part b) , right-hand side). Part c) (bottom-left) illustrates how we conduct the comparison at the heart of our significance testing procedure, in order to test our hypothesis : mainstream news stories that are disproportionately co-shared with fake news consist of a significantly higher presence of misinformation narratives.	141
-----	---	-----

5.2	The top articles from The New York Times (NYT) and The Washington Post (WAPo) in the set of articles disproportionately co-shared with fake news, based on the co-sharing likelihood derived from Twitter sharing behavior. As motivation for our hypothesis in this work, consider the third NYT article and the seventh WAPo article shown in the figure about mail-in ballots and absentee voting. These articles are popular in the same Twitter circles sharing stories published by fake news outlets containing narratives (extracted in this work) that include ‘mail ballot raise risk fraud’, ‘lot people cheat mail’, ‘mail balloting increase incident fraud’, ‘paper mail balloting mean dramatic increase fraud’, and ‘republican complain potential mail absentee ballot fraud’. While only headlines are shown here for illustrative purposes, our investigation also utilized the entire text of the news article. We provide the texts as well as the exact article URLs for all groups of news articles as part of our codebase.	145
5.3	A brief summary of the process of extracting narratives using the method from Ash et al. (2021). The following example sentence, obtained from a text of a fake news article in our dataset, is used: “Over 100,000 U.S. children are maimed or killed by vaccines each year.”	151
5.4	Complete flowchart of the exact implementation of narrative extraction in the open-source Python package: RELATIO (https://github.com/relatio-nlp/relatio); the image is copied directly from Ash et al. (2021).	152

6.1	The objective of the new topic modeling-based method we plan to propose in future research is to help humans (who are studying a particular discourse) find distinctive topics that are present in one but not the other collection; where authors of the two collections are asymmetrically related. C_1 and C_2 represent the two text collections.	191
A.1	Runtime comparison for SCHOLAR) and our own SCHOLAR+ BAT - Note that the overhead due to BAT is only a due to the training and inference time required to obtain the DISTILBERT encoder logits on the full dataset first, and once the teacher logits are available, the run time of both models is the same. We depict the full approximate time (in hours) including this initial overhead in case of BAT.	220
A.2	Modified portions of W-LDA model to accommodate BAT. We omit definitions of additional command-line arguments and data loading, but they are similarly brief.	221
A.3	Instructions for (a) word intrusion and (b) ratings	231
A.4	Descriptions for (a) NYTimes and (b) Wikipedia.	231
B.1	Training set results when modeling the entire dataset and predicting legislators' issue-attention. The corresponding held-out set results are shown in Fig. 3.5a.	238
B.2	Training set results when modeling each congressional session separately and predicting legislators' issue-attention. The corresponding held-out set results are shown in Fig. 3.5b.	238

B.3	Held-out set results with error bars when modeling each congressional session separately and predicting legislators' issue-attention. These correspond to the held-out set results (without error bars) shown in Fig. 3.5b.	239
B.4	Comparison for all legislator attributes after removing non-substantial non-policy issues. We find the same pattern as when using all topics: PAC attribute explains legislators' issue-attention the most.	240
B.5	Held-out set results when issue-attention is quantified using the outputs of an entirely different topic modeling method (as opposed to LDA): the Structured Topic Model (STM).	241
B.6	Held-out set results, congress-by-congress, including results for the additional attribute of legislator <i>seniority</i>	242
B.7	Comparing predictive capacity for issue-attention when considering business or labor PACs in isolation on the held-out set. Business PACs offer significantly higher predictive capacity than labor PACs. However, our significance test reveals that considering all PACs is still significantly better than considering only business PACs.	244
B.8	Specific comparison for the <i>PAC</i> and <i>Committee</i> legislator attributes (cross-validation results). While <i>PAC</i> offers more explanation for issue-attention, results for the combined attribute set of <i>PAC</i> and <i>Committee</i> suggest some complementary information present in these two explanatory variables.	244
B.9	Specific comparison for the <i>PAC</i> and <i>Committee</i> legislator attributes — when modeling each congressional session separately (cross-validation results).	245

B.10	Example of how the rating task looks like to an expert annotator - given their own topic label, and 10 PACs, an annotator rated the association of the PAC with the topic on a 1 – 3 Likert scale.	246
B.11	An example of the kind of real-world relationships that the out- puts of our modeling and analysis can capture automatically - for a PAC (BAE Systems) and an issue (International Security/Foreign Policy) that our model (as well as experts) deem related to one an- other, we find that in a particular congressional cycle (1995-96), there is a significant temporal connection between this PAC and the is- sue, i.e, this PAC donates a significant amount closer to speeches on the particular issue of international security and foreign policy in 2015-16 than for speeches that are not on this issue. In partic- ular, the legislator - Norm Dicks (D-WA) - emerged as a frequent recipient of such temporally significant donations by this PAC, and a simple search on a search engine reveals that this connection is backed by what is known to journalists and validated by real-world knowledge in a straightforward manner. The snapshot in the image is taken from https://armscontrolcenter.org/norm-dicks-the-next-likely-chairman-of-the-house-defense-appropriations-subcommittee/	248

B.12	An example of the kind of real-world relationships that the outputs of our modeling and analysis can capture automatically - for a PAC (National Assn of Federally Insured Credit Unions) and an issue (Finance/Financial regulation) that our model (as well as experts) deem related to one another, we find that in two different congressional cycles (2009-10 and 2015-16), there is a significant temporal connection between this PAC and the issue, i.e, this PAC donates a significant amount closer to speeches on the particular issue of education in 2015-16 than for speeches that are not on this issue. In particular, the legislator - Ed Royce (R - CA) - emerged as the recipient of such temporally significant donations by this PAC in both cycles, and a simple search on a search engine reveals that this connection is backed by what is known to journalists and validated by real-world knowledge in a straightforward manner. Snapshot taken from https://www.americanbanker.com/creditunions/news/big-questions-for-credit-unions-as-royce-retires-from-house	250
C.1	Independent-level shift in member rankings from vote-based to text-based (floor speeches and tweets) ideal points, 115 th and 116 th Congress.	258
D.1	We plot the number of articles in co-shared and control groups against the political partisanship score of the outlet (derived from party affiliation of the Twitter audience sharing political content from the outlets by prior work in Green et al. (2021)).	262

Chapter 1: Introduction

“There are always many different possibilities in language, but choice amongst them is socially structured.”

— Norman Fairclough

“(Choice) is the heart and soul of communicating with language.”

— James Paul Gee

Language both shapes and is shaped by social reality, society, and culture (Downes and Downes, 1998; Barker and Galasinski, 2001; Fairclough, 2013; Gregory and Carroll, 2018); indeed, from a discourse analysis perspective, the analytical framework in Fairclough (1993) is based on the understanding that “on the one hand, processes of text production and interpretation are shaped by (and help shape) the nature of social practice, and on the other hand the production process shapes (and leaves ‘traces’ in) the text, and the interpretative process operates upon ‘cues’ in the text.”¹ Computational analysis of content viewed in terms of the *communicative choices* made in said content can therefore help us better understand the social dynamics of our world and ask new questions as well investigate existing hypotheses in the social sciences.

¹In this particular description, the term *text* is used in the more general sense of *content* and not *form* or written representation.

NLP methods can help realize this potential of computational analysis by working with large unstructured text data, as available data and the size of textual datasets increases rapidly in the digital age (Lazer and Radford, 2017; Ghani et al., 2019; Gupta and Rani, 2019; Sharir et al., 2020; Bender et al., 2021). Unsupervised methods can help directly work with free-form texts since texts may not contain accompanying labels, and existing labels may not fit an intended application or particular research questions behind the use and study of those texts.

Text data is also often available from different sources and venues of discourse. The source as well as other available non-textual contextualizing information (metadata) can be used in addition to the text for answering questions situated in disciplines outside computer science — especially in social and political science. Scientists in these other fields can work with the outputs of automated NLP methods in order to validate and use them for enabling new knowledge; indeed, textual analysis and the concept of text-as-data has been at the heart of some important breakthroughs in social and political science research over the years (Grimmer and Stewart, 2013; DiMaggio, 2015; Wilkerson and Casas, 2017; van Atteveldt and Peng, 2018; Gentzkow et al., 2019; Card, 2019; Müller-Hansen et al., 2020; Maerz and Puschmann, 2020; Dugoua et al., 2022; Grimmer et al., 2022).²

From a *methods* perspective, centering the needs of social science researchers helps ensure that the development of new methods is properly grounded in concrete,

²The allure of the idea of *text-as-data*, an interdisciplinary concept this dissertation is situated in, is captured by Card (2019) as: “... there is an enormous amount of unstructured text available that has been written by people; as with the many other electronic traces we leave in our lives, this sort of ‘found data’ has the potential to be a remarkable source of insight into society, in some cases with the ability to answer questions that have little to do with the original reasons for the existence of such data.” A conference called ‘The New Directions in Analyzing Text as Data (TADA)’ has emerged in recent years to provide a forum “for research on the study of politics, society, and culture through computational analysis of documents” (see: <https://textasdata.github.io/events/>).

Type of Communicative Choice	Brief Description	Relevant Dissertation Chapters
Topics	Themes or categories in text documents (as discovered automatically by topic models)	Ch. 2, Ch. 3, Ch. 4
Labeled Issues	Curated or interpreted topic modeling output to obtain issue labels from domain experts	Ch. 3, Ch. 4
Ideological Issue Frames	Sub-topics or frames used within a political issue by actors of different ideological stripes	Ch. 4
Narratives	Recurring sets of entities and relationships between them ('who did what to whom')	Ch. 5

Table 1.1: The types of communicative choices explored in this dissertation.

useful applications and evaluated with particular goals and use cases in mind (and avoids the tendency to chase the somewhat arbitrary goal of demonstrating ‘state-of-the-art’, especially if the evaluation is disconnected from a method’s use).

From a substantive *empirical research* perspective, viewing text data as a digital trace capturing social behavior and societal dynamics helps recruit NLP methods in order to investigate social phenomena. These investigations and new insights are sometimes possible only when communicative choices made in texts are highlighted and analyzed; the associated metadata, network data, and other types of non-textual data may not always be sufficient on their own for novel sociopolitical analyses. Study designs that can use textual data effectively (via NLP methods) in conjunction with non-textual data (such as associated metadata about the authors) can enable new scientific insights not possible when making use of just one or the other type of data.

We use this understanding to drive much of the research in this dissertation.³ We examine communicative choices in text data, with the types of choices studied shown in Table 1.1), and use associated metadata about authors and contexts of

³The academic ‘we’ is used throughout this dissertation as a deliberate communicative choice made to convey the decidedly collaborative nature of all research discussed in this manuscript.

Type of Contribution	Relevant Dissertation Chapters
New method	Ch. 2.1
Novel, critical validation of existing methods	Ch. 2.2, Ch. 4
New findings in social and political science	Ch. 3, Ch. 4, Ch. 5
New data	Ch. 3, Ch. 4, Ch. 5

Table 1.2: Summary of the various broad types of contributions made in this dissertation. *Validation of existing methods* includes critical examinations of existing methods or evaluation paradigms in methods research as well as assessment of outputs of automated unsupervised methods in order to understand their reliability. *New data* also includes new data generated from outputs of our research, which provides new labels and enables new studies stemming from our findings and their context (for example, grouping mainstream ‘real’ news articles into categories that signify if they share synergy with ‘fake’ news).

the communication such as the media of the text production along with available institutional data. This facilitates an examination of different motivations, context-based variation, and consequences of communicative choices.

1.1 Contributions

In this dissertation, we make several different types of contributions (Table 1.2).

- We introduce a new topic modeling method that can use the knowledge present in large language models such as BERT (Devlin et al., 2019), highlighting some of the advantages that particular neural architectures might offer in topic modeling (Section 2.1).
- We identify critical gaps in a dominant automated evaluation paradigm used to evaluate and compare topic models — in fact, we find that the chosen automated criterion finds differences in the aggregate topic quality of models when humans do not (Section 2.2). Topic modeling literature in the past

few years has defaulted to using one particular automated metric in order to establish the quality of a proposed new topic method and to compare various topic models. Our work, therefore, raises concerns about prior findings on how various topic modeling methods stack up against each other. ⁴

- We find new evidence for the impact donors have on US legislators’ actions using the understudied connection between money and language use. While much of the research on the influence of donors on congressional behavior has looked at legislators’ votes on bills, we show that their donor profile is considerably associated with which issues elected members in the US House of Representatives choose to prioritize in their floor speeches (*i.e.*, their issue-attention). We also contribute a new, comprehensive dataset in order to conduct this novel examination (Chapter 3).
- After validating the outputs of an existing automated text-based ideal point extraction tool, we shed new light on variation in ideological positioning (quantified by the communicative choices of ideological framing and polarized terms used within political issues) based on venue and audience-based considerations. This work contributes to and expands legislative behavior theory in political science. We also construct an original dataset combining various sources of information about members of Congress with three different ideal point measurements, derived from their voting records on bills in Congress, their floor speeches, and their tweets (Chapter 4).

⁴Follow-up work in [Hoyle et al. \(2022\)](#) shows that neural topic models, the new and popular class of topic models being put forth by developers in NLP and machine learning research, not only might be giving a false impression of improvement (as per our contribution detailed in Section 2.2), but fares worse than classical models such as LDA ([Blei et al., 2003](#); [Griffiths, 2002](#)) on evaluation criteria grounded in qualitative content analysis (the dominant use case for topic models outside CS research).

- We contribute a new understanding of mainstream media’s role in online misinformation, by finding that a significantly higher presence of misinformation narratives in particular mainstream news articles is associated with their disproportionate uptake by conveyors of fake news on Twitter. Our quantitative results for testing our hypothesis, as well as a comprehensive qualitative case study, suggest a strategic repurposing of mainstream news by conveyors of misinformation. This further suggests the possibility of the use of mainstream news to enhance the reach and persuasiveness of misleading narratives. Our result has important implications for responsible journalistic practice and our work also represents a significant paradigmatic difference from most current research on misinformation, which focuses on false news reports. The modeling outputs of our work create new categorizations of existing mainstream news articles. into those that are meaningfully adjacent to fake news and those that are not. This new data will enable further research into mainstream media’s potential role in misinformation spread on social media (Chapter 5).

1.2 Roadmap for this Dissertation

In **Chapter 2**, we detail our contributions to topic modeling development and evaluation, explaining a new neural topic model we propose (Section 2.1) before taking a step back to take a critical look at the dominant paradigm of automatically evaluating and comparing topic models in the literature (Section 2.2). In the latter section, we find that the automated criterion finds significant differences between model outputs when humans do not. Our findings challenge the conventional wisdom built into topic modeling development literature in recent years regarding

the comparison between topic modeling quality of neural and classical methods.

Next, in **Chapter 3**, we highlight how unsupervised NLP can be used to study motivations behind communicative choices. In particular, we show that information about a US House representative’s donors is associated with how they distribute their attention across different policy issues in their floor speeches. This association is meaningful: the association is higher than other relevant sources of information about the representative, the pattern holds consistently over a time period of twenty-four years, and automatically learned donor-issue associations are deemed meaningful by experts. Looking at associations between language use and communicative choices helps provide a new avenue for looking at the influence of money in politics.

Chapter 4 uses computational modeling to investigate how different institutional constraints and audience considerations help explain the variation in political actors’ communicative choices and ideological positioning.

In **Chapter 5**, using social network data to complement text data, we use another unsupervised NLP method to study the consequences of communicative choices by examining mainstream media articles that are popular in fake news circles on Twitter. We find that a significantly higher presence of existing misinformation narratives characterizes such mainstream news articles.

Chapter 6 concludes this dissertation work and offers a robust research agenda for future work, some of which is based on our prior research efforts and preliminary findings that can both concretely motivate and guide proposed future research.

Note on the background literature for this dissertation. We provide relevant background and prior work in each individual chapter of this dissertation.

Chapter 2: Extracting Communicative Choices from Textual Content: Enhancing and Comparing Topic Models

Topic modeling is one of the most popularly adopted NLP methods outside of computer science-related disciplines for a variety of applications (Boyd-Graber et al., 2017).¹

In brief, topic modeling refers to a family of typically unsupervised methods that takes as input a collection of documents and models them as a collection of latent categories or topics. Let the document collection contain N documents and V terms in the vocabulary, and let the number of topics be K (decided by the user in case of parametric models, though non-parametric extensions can try to optimize for and select K themselves). Probabilistic topic models such as LDA (Blei et al., 2003) represent each topic k as a probability distribution or mixture over the terms in the vocabulary (β_k ; each row is part of the $K \times V$ dimensional topic-word matrix $\mathbf{B}$), and represent each document d similarly as a distribution over the K discovered topics (θ_d ; each row is part of the $N \times K$ dimensional document-topic matrix $\boldsymbol{\theta}$). Automatically discovered topics, or the latent dimensions, can be *interpreted* using the rankings they produce over the terms in the vocabulary (using $\mathbf{B}$) and over the documents in the collection (using $\boldsymbol{\theta}$).

¹Indeed, Fig. 2.1 provides some informal evidence for this notion using publications outside computer science citing a topic modeling method (LDA) and publications citing the popular NLP language modeling method of BERT. The figure is adapted from the presentation of the joint work in Hoyle et al. (2021).

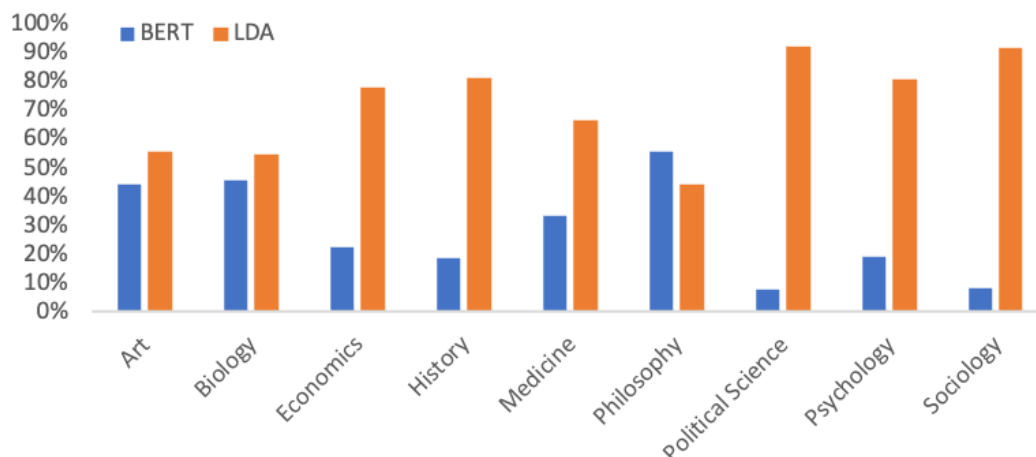


Figure 2.1: We compare the usage of BERT and LDA models to emphasize the uptake and impact of topic modeling methods in research conducted outside the field of computer science. This comparison is done using research works between 2019 and 2021 (using just one topic modeling method: LDA). The data on citations and research areas of the publications come directly from Semantic Scholar (<https://www.semanticscholar.org/>), and the % of publications extracted in each area that cite BERT (Devlin et al., 2019) or LDA (Blei et al., 2003) are shown.

We considered topics to be a form of *communicative choice*, and topic modeling as a method that can extract these salient choices from text.

Roadmap for this chapter. In this chapter, we will first introduce a method based on knowledge distillation (Hinton et al., 2015) that incorporates contextual linguistic knowledge as a modular addition to virtually any neural topic model (NTM). This helps alleviate a crucial shortcoming in most topic models while also showcasing a novel application for the dominant *pretrain and fine-tune* paradigm that uses large transformer-based language models in NLP (Section 2.1).

Next, in Section 2.2, we conduct a close examination of the near-universal automated evaluation paradigm in topic modeling method development research in NLP and machine learning. This paradigm uses one particular metric (and its variants) to compare different topic models and claim a new “state-of-the-art” in

topic modeling. We find gaps in this automated evaluation framework that call into question much of recent topic modeling work claiming NTMs as ‘better’ than LDA.²

Our work in [Hoyle et al. \(2022\)](#) represents an important follow-up to the work discussed in Section 2.2; it operationalizes different evaluation criteria that are grounded in qualitative content analysis — the dominant use case of topic modeling application outside method development. Together, the results in Section 2.2 in this chapter and in [Hoyle et al. \(2022\)](#) challenge the tide of results claiming NTM as ‘better’ than the previous non-neural method of choice. Indeed, in the latter work, we find that NTM methods fare worse than the previous non-neural method of choice (LDA) on evaluation criteria grounded in the use case of qualitative content analysis. Our work emphasizes the importance of validating the outputs of NLP methods as well as grounding their evaluation in well-defined use cases.

2.1 Incorporating Contextual Knowledge in Neural Topic Models

While NTMs can take various forms, including variations on the underlying distributions used for the latent variables and the exact form of the autoencoder architecture used, they tend to treat input documents as a non-sequential collection of terms or bag-of-words (much like LDA). The sequential nature of texts in the English language (as a linearization of the hierarchical structure of language) is one of its more important properties, and as such, the NLP community has made advances in sequential language modeling ([Devlin et al., 2019](#)) using transformer architectures ([Vaswani et al., 2017](#)). The typical framework for utilizing these transformer-based language models is generally called *pre-train and fine-tune*.

Briefly, a language model with a large number of parameters, trained on large

²Portions of the work in this chapter are collaborative; specific details will be provided in each of the two sections.

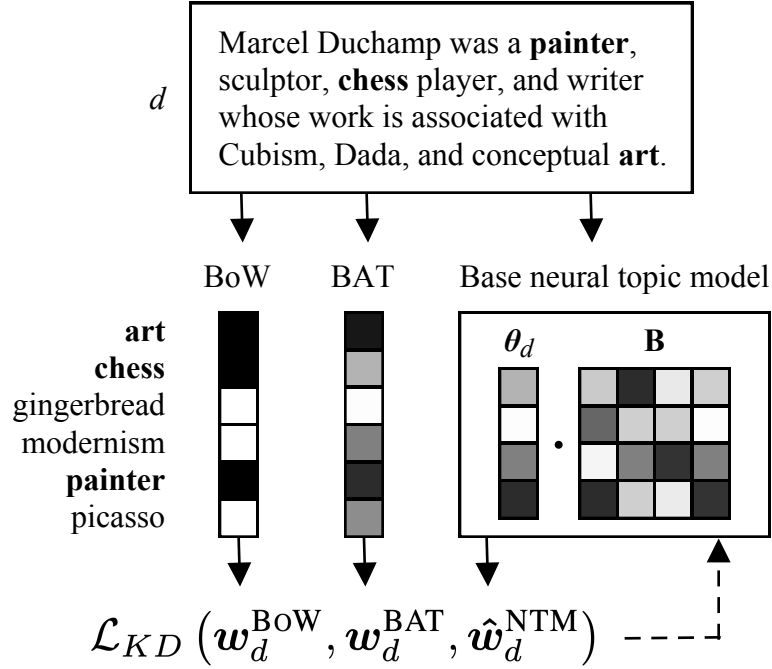


Figure 2.2: Improving a base neural topic model with knowledge distillation. A document is mapped through both a standard BoW representation and a **B**ERT-based Auto-encoder “Teacher” (BAT), yielding two distributions over words. These are used as the ground truth in the “student” topic model’s document reconstruction loss $\mathcal{L}_{KD}$ (backpropagated along the dotted line). Crucially, the BAT distribution assigns mass to unobserved but related terms (unbolded).

quantities of text data in an unsupervised manner, is used as a *pretrained model* to represent texts in a collection. The unsupervised language modeling objective consists of variations on predicting ‘masked’ words from the surrounding context. These representations are then *fine-tuned*, *i.e.*, adjusted to do well (as defined by an evaluation criterion) on a particular objective as part of a particular computational task.³ The relative success of these models (on particular NLP tasks) is based, in part, on their ability to capture linguistic context, usually modeling the position of words in a text explicitly and therefore not treating texts as simply a bag-of-words during their training (Clark et al., 2019; Liu et al., 2019b; Rogers et al., 2020).

³Raffel et al. (2020) provide a detailed overview of the pre-training and fine-tuning process and how it has been conceived and applied especially in NLP.

However, there is a limitation to these models and the typical application of pre-train and fine-tune paradigm: in having a huge number of parameters in order to gain its contextual prowess from large quantities of data, these models often become uninterpretable and opaque (‘black box’). Topic models, meanwhile, derive their interpretability from a dramatic dimensionality reduction that preserves words and documents as readable insights into inferred distributions.

The core innovation of our method is presented in Fig. 2.2.⁴ Our method combines the interpretable method of understanding underlying themes in a collection through topic models and the context-capturing abilities of large pretrained language models using knowledge distillation (Hinton et al., 2015), creating a best-of-both-worlds scenario.⁵

Roadmap for this section. We first describe our proposed method, providing an overview of knowledge distillation and how it can be used with the pretrained BERT language model (Devlin et al., 2019). Our method modifies the inference procedure of a typical autoencoder-based NTM (Section 2.1.1).

⁴Contents of Fig. 2.2, including the description, are directly taken from Hoyle et al. (2020).

⁵The description of the work in Section 2.1 is adapted from joint work done in: **Pranav Goel***, Alexander Hoyle*, and Philip Resnik. “Improving Neural Topic Models Using Knowledge Distillation.” Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020. (* indicates equal contribution).

Goel’s primary contributions to this work include:

- Jointly with his colleague (Hoyle), coming up with the methodological innovation on how the two unsupervised approaches of fine-tuning language models and topic modeling could be combined.
- Jointly with his colleague (Hoyle), setting up and implementing the experimental baselines.
- Creating and implementing the topic alignment procedure to study the effect of the proposed method on the level of individual topics (Section 2.1.4).
- Understanding and explicating prior literature on automated evaluation measures that help compare different topic modeling methods, and more broadly, contextualizing and situating our efforts in the literature.
- Helping identify relevant examples and creating figures used to convey our findings.

We then explain the evaluation method and paradigm that follows an established norm in the neural topic modeling literature in Section 2.1.2 (which will be challenged in the next part of this chapter in Section 2.2). Details on the datasets, baseline topic models, and other experimentation specifics along with the results are provided in Section 2.1.3.

While most previous literature describes the *aggregate* results alone, we further analyze what changes our modular approach introduces at the level of *individual* topics when added to a base NTM. We do this by adapting a bipartite graph matching procedure in order to compare topics from two different topic models for the first time — this procedure is a general contribution to topic modeling analysis (Section 2.1.4). We summarize our findings and the key takeaways from the development of this novel topic modeling method in Section 2.1.5, explicate some related work and background in Section 2.1.6, and then take a step back to critically analyze the automated topic modeling evaluation procedure in the next portion of our work (Section 2.2).

2.1.1 Method

Inference process in neural topic models. Although our method is not dependent on the choice of the underlying NTM, we use the work and the SCHOLAR model introduced in Card et al. (2018) for our formal description below as well as our base implementation. We note that SCHOLAR generalizes one of the first VAE-based NTMs in PRODLDA (Srivastava and Sutton, 2017). The generative story is similar to LDA. One of the primary differences in this instantiation is that the uniform Dirichlet prior is replaced with a logistic normal (LOGNORMAL):⁶

⁶This choice was originally made due to the location-scale limitation for the reparameterization trick in VAEs. However, recent developments (e.g., Figurnov et al., 2018) have helped move the field

For each document d :

- Draw topic distribution $\boldsymbol{\theta}_d \sim \text{LOGNORMAL}(\boldsymbol{\alpha}_0)$
- For each word w_{id} in the document:

$$w_{id} \sim \text{Multinomial}(1, f(\boldsymbol{\theta}_d, \mathbf{B}))$$

Following PRODLDA (Srivastava and Sutton, 2017), $\mathbf{B}$ is a $K \times V$ matrix where each row corresponds to the k th topic-word probabilities (in log-frequency space). The multinomial distribution over a document’s words is parameterized by

$$f(\boldsymbol{\theta}_d, \mathbf{B}) = \sigma(\mathbf{m} + \boldsymbol{\theta}_d^\top \mathbf{B}) \quad (2.1)$$

where $\mathbf{m}$ is a vector of fixed empirical background word frequencies and $\sigma(\cdot)$ is the softmax function. We note that each document is treated as a bag of words ($\mathbf{w}^{\text{BOW}_d}$).

VAE-based models such as SCHOLAR perform inference via an approximation of the true intractable posterior $p(\boldsymbol{\theta}_d | \cdot)$ using a neural *encoder* network $g(\mathbf{w}_d)$, that is used to parameterize the variational distribution $q(\boldsymbol{\theta}_d | g(\cdot))$ (here, a logistic normal with diagonal covariance). The Evidence Lower BOund (ELBO) is therefore

$$\begin{aligned} \text{ELBO} &= \mathbb{E}_{q(\boldsymbol{\theta}_d | \cdot)}[\mathcal{L}_R] \\ &\quad - \text{KL}[q(\boldsymbol{\theta}_d | \mathbf{w}^{\text{BOW}_d}, \mathbf{x}_d) || p(\boldsymbol{\theta}_d)], \end{aligned} \quad (2.2)$$

$$\mathcal{L}_R = (\mathbf{w}_d^{\text{BOW}})^\top \log f(\boldsymbol{\theta}_d, \mathbf{B}), \quad (2.3)$$

which is optimized using stochastic gradient descent. This particular form of the reconstruction error $\mathcal{L}_R$ is a result of independent multinomial draws.⁷

beyond that particular restriction, and one of the models used in this work —Burkhardt and Kramer (2019) — demonstrates this by positing several Dirichlet-based NTMs that use VAEs.

⁷The formal description of the inference procedure, along with Eqs. (2.1) to (2.3), is taken directly from Hoyle et al. (2020).

Knowledge distillation. *Knowledge distillation* (KD, [Hinton et al., 2015](#)) is a procedure relying on one basic framework: a parameter-rich classifier trained on large amounts of data, called a *teacher* model, is used to guide a smaller *student* model. The guidance is based on high-quality probability estimates over outputs in the teacher model.

Our work relies on the fact that the knowledge learned by a teacher model (which can be a deep neural network) is contained in its probability estimates over classes (and not just its parameters). Therefore, given any particular input, the teacher estimates can be reused as ‘soft’ labels in order to train the student model.

The basic intuition is that since the information contained in these estimates is useful, the student should require less training data and can achieve better generalization. For example, a text discussing a baseball match will result in higher label probabilities for the topic of cricket than the topic of mass shootings. In practice, the information present in teacher model probabilities is combined with the signal present in true labels (at training time) via linear interpolation, as presented below for our approach (Eq. (2.4)).

Using knowledge distillation: Unsupervised large language models as the ‘teacher’ for the topic model autoencoder ‘student’. Knowledge distillation (KD) is typically used in *supervised* settings ([Gou et al., 2021](#)). In this work, we apply KD to *unsupervised* topic modeling, which has not previously been attempted to the best of our knowledge. In order to do this, we use the fact that both an autoencoder and a topic model share the same unsupervised “task”: reconstructing the original document, which formally implies predicting a distribution over the vocabulary (derived from a collection of input documents). The BERT autoencoder serves as our teacher model and offers a dense prediction based on its encoding of rich sequential

information (as a result of pretraining on a large corpus). We use it to guide the topic model (as the student), which is generating its own prediction of that distribution.

Specifically, consider Eq. (2.1) as a latent logistic regression that aims to approximate the distribution over words in a document. This view ties directly to the typical supervised learning setup used in KD. With each word viewed as a label, the NTM’s objective of predicting the word distribution can be viewed as a multi-label classification. We thus obtain a very standard format: a softmax over logits estimated by a neural network, and a cross-entropy loss utilized for training the model.

However, an important difference with the typical multi-label classification setup is the extreme sparsity of the bag-of-words distribution ($\mathbf{w}_d^{\text{BOW}}$). This limits the model’s ability to generalize as during backpropagation (Eq. (2.3)), the update to topic parameters will only account for observed terms, which can result in overfitting and less coherent topics.

In direct contrast to this sparsity of the bag-of-words representation and the limitation of accounting for only the observed terms, dense document representations generated by large language models like BERT can capture rich, contextual information. This is the concrete motivation for our approach, which we call **BERT-based Autoencoder as Teacher (BAT)**. Our method includes the following components:

- **The knowledge distillation objective:** To apply KD to a “base” neural topic model, we replace the reconstruction term $\mathcal{L}_R$ in Eq. (2.3) with $\mathcal{L}_{KD}$, as follows:

$$\begin{aligned}\mathbf{w}_d^{\text{BAT}} &= \sigma\left(\mathbf{z}_d^{\text{BAT}}/T\right) N_d \\ \hat{\mathbf{w}} &= f(\boldsymbol{\theta}_d, \mathbf{B}; T) \\ \mathcal{L}_{KD} &= \lambda T^2 \left(\mathbf{w}_d^{\text{BAT}}\right)^\top \log \hat{\mathbf{w}} + (1 - \lambda) \mathcal{L}_R\end{aligned}\tag{2.4}$$

Here, $\mathbf{z}_d^{\text{BAT}}$ are the logits produced by the teacher network for a given input document d , meaning that $\mathbf{w}_d^{\text{BAT}}$ acts as a smoothed pseudo-document. T is the softmax temperature, which controls how diffuse the estimated probability mass is over the words (hence $f(\cdot; T)$ is Eq. (2.1) with the corresponding scaling). This differs from the original KD in two ways: (a) it scales the estimated probabilities by the document length N_d , and (b) it uses a multi-label loss.⁸

- **The teacher model:** The teacher logits $\mathbf{z}^{\text{BAT}}$ are generated using the pretrained transformer DISTILBERT (Sanh et al., 2019), which is itself a distilled version of BERT (Devlin et al., 2019).⁹ BERT-like models are generally pretrained on large corpora (that cuts across domains) using an objective similar to traditional language modeling. This allows these models to capture linguistic context more effectively than bag-of-words models (Clark et al., 2019; Liu et al., 2019b; Rogers et al., 2020). Recall the aforementioned variational auto-encoder-based formulation for NTMs. Following from that, we treat DISTILBERT as a *deterministic* auto-encoder and fine-tune it with the document-reconstruction objective $\mathcal{L}_R$ (Eq. (2.3)) on the same dataset. Thus, we use a BERT-based Autoencoder as our Teacher model (BAT).¹⁰

We described the **student model** in our discussion of the inference process above for NTMs (Eqs. (2.1) to (2.3)).

⁸We note that the formal description of our knowledge distillation objective as well as Eq. (2.4) are taken directly from the description in Hoyle et al. (2020).

⁹We use DISTILBERT since its lighter weight can accommodate longer documents, which is especially important for topic modeling. However, even then we are forced to divide longer documents into chunks, estimate logits for each chunk, and finally take the pointwise mean. In our preliminary experiments, more complex schemes (i.e., LSTMs, Hochreiter and Schmidhuber, 1997) did not provide any benefit.

¹⁰Since the completion of this work, Bianchi et al. (2021) has provided a path for models like BERT to offer document representations as topic modeling input in their proposed contextual topic model (CTM).

- **Clipping the logit distribution:** Depending on preprocessing, the vocabulary V can contain tens of thousands of terms or tokens. This results in a long tail of probability mass assigned to unlikely terms, which violates standard assumptions of sparsity. [Tang et al. \(2020\)](#) empirically demonstrate that truncating the logits to the top- n classes (and assigning uniform mass to the rest) improves accuracy (in a classification setting). Inspired by that finding, we choose the top $c N_d, c \in \mathbb{R}^+$ logits and assign *zero* probability to the remaining classes (vocabulary items) in order to enforce sparsity.

2.1.2 Coherence-based Automated Evaluation

The central goal of topic models is obtaining human-understandable or interpretable categories. *Coherence* is one way that the ‘goodness’ of interpretability of topics has been conceptualized. The term has its roots in Latin *cohaerere*, which means “to stick together.” Coherence, from a human reader’s perspective, can be thought of intangibly as a set of terms or items, which when viewed together enable human recognition of an identifiable category. This understanding of topic coherence aligns with Propositions 2 and 3 of [Doogan and Buntine \(2021\)](#): “an interpretable topic is one that can be easily labeled” (by humans), and “has high agreement on labels” (when different human readers independently try to label a topic discovered by a topic model). Much like the concept of *interpretability*, that of real-world *coherence* is “simultaneously important and slippery” ([Lipton, 2018](#)).

Human judgments of topic quality, interpretability, and coherence have been obtained in different ways ([Newman et al., 2010](#); [Mimno et al., 2011](#); [Aletras and Stevenson, 2013](#); [Chang et al., 2009](#)). We will cover two of those ways in Section 2.2.

[Chang et al. \(2009\)](#) found that document perplexity, the automated metric of

choice for measuring topic modeling outputs till that point in the field, was *negatively* correlated with aggregated human judgments of topics discovered by a topic model. Subsequently, attempts in the topic modeling literature to automatically measure topic quality that positively and significantly correlates with human judgment settled on ways to operationalize coherence as an automated metric (Newman et al., 2010; Aletras and Stevenson, 2013; Lau et al., 2014). As a result of these efforts to validate automated measures, the field has adopted pairwise normalized pointwise mutual information (NPMI) as the metric of choice for evaluating and comparing two different topic models (Aletras and Stevenson, 2013; Lau et al., 2014). This is also backed by the results of our meta-analysis of the neural topic modeling literature discussed in Section 2.2.1.¹¹

Here we use NPMI to automatically evaluate topic modeling quality.¹²

NPMI scores topics highly if the top N words—summed over all pairs w_i and w_j —have high joint probability $P(w_j, w_i)$ compared to their marginal probability.¹³

$$\sum_{j=2}^N \sum_{i=1}^{j-1} \frac{\log \frac{P(w_j, w_i)}{P(w_i)P(w_j)}}{-\log P(w_i, w_j)}. \quad (2.5)$$

Word co-occurrence counts, derived from a *reference corpus* for a specific context window (which can range from ten words to the entire document), are used to estimate the probabilities. Therefore, the choice of reference corpus is a direct

¹¹We note here that while the human validation of the NPMI metric was only done with non-neural probabilistic methods like LDA, the neural topic modeling literature adopted the metric as a way to compare the proposed neural method with other topic models with no additional human verification — we explore consequences of this in Section 2.2. Returning to this study of evaluating our proposed new topic modeling method in light of those results is a topic for future work.

¹²While the aim of this work is to propose a method that can improve topic coherence, we also show that our knowledge-distillation-based augmentation does not hurt downstream performance on a classification task in Appendix A.2.

¹³Alternative metrics exist, but they typically also rely on either joint probability estimates or NPMI directly (e.g., C_v Röder et al., 2015).

determinant of the strength of human correlation (Lau et al., 2014; Röder et al., 2015).¹⁴ Other decisions involved in calculating the value for this particular metric include the size of the context window size (for obtaining co-occurrence counts) and the number of top terms of the topic (N).

Following precedent, we use the top ten words in each topic ($N = 10$) to calculate the NPMI score for that individual topic. *Internal* NPMI uses a held-out set from the same corpus when estimating the reference co-occurrence counts. While most prior work defaults to internal NPMI as the primary metric, we also provide results based on *external* NPMI where a different corpus (the Gigaword 5 dataset (Parker et al., 2011) in our case) is used as the reference (following Card et al. (2018)).

In order to compare our proposed method with other topic modeling approaches, the NPMI values for individual topics are aggregated by taking the average over all the topics. This single value is then used to compare different topic models (on the same dataset and with the same choice for the number of topics K). This procedure represents the typical automated evaluation and comparison method of choice in much of the topic modeling literature and this is what we follow below in Section 2.1.3. Since our method is a *modular* addition that can be made to any base neural topic model, we further examine how our knowledge-distillation-based addition impacts *individual topics* in Section 2.1.4, providing a crucial addition to the aggregate-only comparison that dominates the literature.

D	V	Avg N_d	Preprocessing details
20NG 18k	2k	87.1	Srivastava and Sutton (2017)
Wiki 28.5k	20k	1395.4	Nan et al. (2019)
IMDb 50k	5k	95.0	Card et al. (2018)

Table 2.1: Corpus statistics, which vary in total number of documents (D), vocabulary size (V), and average document length (N_d).

2.1.3 Experimental Details and Results

Data. We validate our proposed method using three available datasets that span different domains, vocabulary sizes, numbers of documents, and document lengths: 20 Newsgroups (20NG, [Lang, 1995](#)),¹⁵ Wikitext-103 (Wiki, [Merity et al., 2017](#)),¹⁶ and IMDb movie reviews (IMDb, [Maas et al., 2011](#)).¹⁷ These are commonly used in neural topic modeling using the preprocessed versions provided by various authors for benchmarking topic models (details are provided in the references included in Table 2.1).¹⁸ Following prior work, we use the following train/dev/test splits: 48/12/40 for 20NG, 70/15/15 for Wiki, and 50/25/25 for IMDb.¹⁹

Experimental Baselines. In order to provide a robust measure of improvement, we use three experimental baseline models that span different neural topic modeling

¹⁴In using NPMI, we follow the topic modeling literature in machine learning and NLP communities seeking to develop and propose new topic models. We also use the precedent in this specific literature to guide our other decisions for comparing our topic model with relevant baselines. However, our experience here with the use of aggregated automated judgments alone informs and motivates the next part of this chapter, which examines NPMI and the automated topic modeling evaluation paradigm critically (Section 2.2).

¹⁵qwone.com/~jason/20Newsgroups

¹⁶s3.amazonaws.com/research.metamind.io/wikitext/wikitext-103-v1.zip

¹⁷ai.stanford.edu/~amaas/data/sentiment

¹⁸Contents of Table 2.1, including the description, are directly taken from [Hoyle et al. \(2020\)](#).

¹⁹The splits are used to estimate NPMI — dev splits are used to select hyperparameters, and test splits are run after hyperparameters are selected and frozen.

styles.²⁰ All these baseline models were shown to achieve the highest NPMI on the majority of the respective datasets used in corresponding works and provided a considerable improvement over previous neural and non-neural topic models (such as [Srivastava and Sutton, 2017](#); [Miao et al., 2016](#); [Ding et al., 2018](#)). Prior to our work here, these models have not been compared head-to-head due to the contemporaneous nature of these works.

SCHOLAR. [Card et al. \(2018\)](#) use a VAE-based ([Kingma and Welling, 2014](#)) NTM setup (as introduced in [Srivastava and Sutton, 2017](#)) using a logistic normal prior (Section 2.1.1), and contribute a sophisticated method to incorporate the metadata associated with the documents.

DVAE. [Burkhardt and Kramer \(2019\)](#) use a Dirichlet prior and follows the same generative story as the original LDA formulations presented in [Blei et al. \(2003\)](#). This model can therefore take advantage of useful properties of the Dirichlet such as multi-modality ([Wallach et al., 2009](#); [Wang et al., 2020a](#)).

W-LDA. [Nan et al. \(2019\)](#) use a Wasserstein auto-encoder ([Tolstikhin et al., 2018](#)) instead of a VAE, and use a Dirichlet prior that is matched by minimizing the maximum mean discrepancy. They demonstrate state-of-the-art coherence on several datasets and also note that topics provided by W-LDA tend to exhibit greater diversity in their top terms.

In order to highlight the modularity of our proposed innovation, we combine our KD-based method with both SCHOLAR and W-LDA in our results below.

Our model settings. As explained in Section 2.1.1, our approach includes a “base” NTM (as the student model) and unnormalized probabilities (logits) over words

²⁰Note that “baseline” is used in the traditional sense of benchmarking model development in machine learning, and is different from the usage of our terminology of “base” NTM that gets augmented by knowledge distillation (Section 2.1.1).

estimated by DISTILBERT, a transformer model, as the teacher.

In this work, we use both SCHOLAR and W-LDA as our base models. SCHOLAR is our primary model and constitutes our point of comparison with the baselines, while we use W-LDA to offer a proof-of-concept for our method’s advantage of modularity — in fact, we note that our implementation of knowledge distillation can be added to W-LDA using only a few lines of code (Appendix A.5).

We evaluate and compare all models at both $K = 50$ and $K = 200$ topics (again following prior NTM development literature). On each dataset, the models are tuned using NPMI calculated using reference co-occurrence counts derived from a held-out development set (a split of the corresponding corpus; internal NPMI). For our baselines, we use the publicly-released implementations.²¹ When available, we use the original hyperparameter settings as our starting point but do perform an exhaustive grid search on the SCHOLAR baselines and SCHOLAR + BAT to ensure a fair comparison. The ranges for the grid search, optimal values we obtain, and other details are provided in Appendix A.4.1, while details on our compute infrastructure and runtimes for models are included in Appendix A.4.2.

We note that our method also introduces some new hyperparameters: weight for KD loss, λ (Eq. (2.4)); softmax temperature T ; and the clipping variable or the proportion of the word-level logits (provided by the teacher model) that gets retained (relative to document length, see clipping in Section 2.1.1). For most pairs of dataset- K , topic quality gets improved under most settings using a relatively small set of values for each hyperparameter, which leads to better results.²²

²¹SCHOLAR: github.com/dallascard/scholar; W-LDA: github.com/awsmlabs/w-lda; DVAE: github.com/sophieburkhardt/dirichlet-vae-topic-models. For our models augmented using knowledge distillation, we utilize our own reimplementations of the baseline methods within a common codebase after validating the reimplementations via comparable results to the reported numbers by the authors on corresponding datasets.

²²In fact, after an extensive search on SCHOLAR + BAT, W-LDA could be tuned within a few iterations.

Since topic models use random sampling procedures, we run each model five times and provide the average values across the runs in order to provide a robust comparison (a step that was not taken or reported by authors of our baseline methods).

We fine-tune a slightly modified version of DISTILBERT (teacher) using the same document reconstruction objective as the NTM or the student model ($\mathcal{L}_R$, Eq. (2.3)) on the training data split. DISTILBERT first maps a WordPiece-tokenized (Wu et al., 2016) document d to an l -dimensional hidden vector using a transformer architecture (Vaswani et al., 2017). Then, it maps the hidden vector back to logits over V words. Following the restrictions for the DISTILBERT model, we split long documents into blocks of 512 tokens and then pool the transformer outputs using the mean. We use the publicly available pretrained DISTILBERT model (Wolf et al., 2020). Keeping the held-out dev set data split consistent across the topic model and the transformer, we train until perplexity converges on the dev set. DISTILBERT achieves dramatically lower perplexity than all topic model baselines, as expected. We note that the teacher transformer model needs to be trained only once per corpus to obtain the logits $\mathbf{z}^{\text{BAT}}$ used to guide the student NTM, and we can then experiment with different NTM variations using those same logits.

	$K = 50$			$K = 200$		
	20NG	Wiki	IMDb	20NG	Wiki	IMDb
DVAE	0.340	0.490	0.145	0.316	0.450	0.160
W-LDA	0.279	0.494	0.136	0.188	0.308	0.095
SCHOLAR	0.322 (0.007)	0.494 (0.005)	0.168 (0.002)	0.263 (0.002)	0.473 (0.005)	0.140 (0.001)
SCH. + BAT	0.354 (0.004)	0.521 (0.009)	0.182 (0.002)	0.332 (0.002)	0.513 (0.001)	0.175 (0.003)

Table 2.2: The NPMI for our baselines (Section 2.1.3) compared with BAT (explained in Section 2.1.1) using SCHOLAR as our base neural architecture. We achieve better NPMI than all baselines across three datasets and $K = 50, K = 200$ topics. We use 5 random restarts and report the standard deviation.

Results. Using the VAE-based SCHOLAR as the base model, our method (BAT) discovers more coherent than the previous state-of-the-art baseline NTMs (Table 2.2) as measured by NPMI.²³ Specifically, we obtain better coherence scores than the DVAE and W-LDA baseline models as well as the baseline of SCHOLAR (without our KD-based augmentation). For a robust comparison, we take the mean of all results obtained across runs with different random seeds and find consistent improvement across all baselines and datasets, as well as a smaller ($K = 50$) and a larger ($K = 200$) number of topics.

In addition to this aggregated improved performance, we highlight once again the modular nature of BAT — it can work with different kinds of base models since it makes very few assumptions about that underlying model. Our approach only requires a word-level reconstruction objective, which holds for the majority of proposed NTMs (to the best of our knowledge). In order to demonstrate this, we use the Wasserstein auto-encoder (W-LDA) as the base NTM (as opposed to the VAE-based SCHOLAR model used earlier), and find that augmenting with BAT improves upon that underlying model that utilizes a different architecture Table 2.3.²⁴ We also provide preliminary results for the DVAE baseline model augmented with our BAT approach in Appendix A.3.

We report the dev set results (corresponding to the test set results in Tables 2.2 and 2.3) in Appendix A.1, where the same pattern of results holds across all settings.

Finally, we provide external NPMI results where NPMI is computed using reference counts from an external corpus (different from the training corpus). Recall that we use (Gigaword 5, [Parker et al., 2011](#)) as the external corpus for SCHOLAR

²³Contents of Table 2.2, including the description, are directly taken from [Hoyle et al. \(2020\)](#).

²⁴We note that for $K = 200$, the W-LDA baseline model did not tune well, in part due to the model’s extensive run time. We focus on augmenting that model for 50 topics, which is consistent with the number of topics used in the work that proposed the model [Nan et al. \(2019\)](#). Contents of Table 2.3, including the description, are directly taken from [Hoyle et al. \(2020\)](#).

and SCHOLAR + BAT — the same patterns hold in all but one setting (Wiki, $K = 50$; Table 2.4).²⁵ BAT improves topic coherence relative to SCHOLAR. Our model’s ability to use generalized language from BERT pretrained on a large, domain-general corpus (as opposed to just bag-of-words from the training corpus) as well as the fact that our fine-tuning procedure does not overfit to the training data together helps explain the external NPMI results we obtain.

	20NG	Wiki	IMDb
W-LDA	0.279 (0.010)	0.494 (0.012)	0.136 (0.008)
+ BAT	0.299 (0.010)	0.505 (0.014)	0.162 (0.003)

Table 2.3: Mean NPMI (s.d.) across 5 runs for W-LDA Nan et al. (2019) and W-LDA + BAT for $K = 50$, showing improvement on two of three datasets. This demonstrates that our method is *modular* and can be used with base neural topic models that vary significantly in architecture.

K		SCHOLAR	+ BAT
50	20ng	0.147 (0.002)	0.170 (0.006)
	Wiki	0.193 (0.006)	0.187 (0.004)
	IMDb	0.149 (0.003)	0.161 (0.003)
200	20ng	0.111 (0.001)	0.171 (0.002)
	Wiki	0.177 (0.003)	0.190 (0.008)
	IMDb	0.122 (0.002)	0.159 (0.003)

Table 2.4: External NPMI (s.d.) for the base SCHOLAR and SCHOLAR + BAT. Models selected according to performance on the development set using internal NPMI.

2.1.4 Topic Alignment based Comparison and Analysis

So far, we have followed precedent and much of standard practice in topic modeling evaluation in order to obtain an aggregated comparison that shows that our model finds more coherent topics (as measured by NPMI) *on average*. Motivated

²⁵Contents of Table 2.4, including the description, are directly taken from Hoyle et al. (2020).

by the modularity of our approach, we now study the impact of BAT on *individual topics* in order to determine if the aggregate improvement meaningfully affects their quality. We note that the procedure we contribute here can be used more generally to compare two topic models on a topic-by-topic basis and facilitate human understanding of the differences and similarities in the latent concepts being discovered by the two models.

First, we align (or match) the topics discovered by the base SCHOLAR model and the ones discovered by the augmented SCHOLAR + BAT model. We use a variation of the competitive linking algorithm that approximates a greedy graph matching procedure to a weighted bipartite graph (Melamed, 2000). We construct a fully connected weighted bipartite graph by linking all topic pairs across (but not within) the two models to get $K \times K$ links. Each link (or edge) is assigned a weight calculated using the Jensen-Shannon (JS) divergence (Wong and You, 1985; Lin, 1991) between the word distributions of the top terms representing the two node topics. Starting with the full set of pairs, the pair (t_i, t_j) with the lowest JS divergence score is added to the alignment (constituting the final output of this procedure). Then the two topics (one from each model) — t_i and t_j — are removed from consideration. From the remaining set, the pair with the lowest JS divergence score is picked again, and we iterate until no pairs are left. We then sort the resulting aligned topic pairs using their JS divergence scores in order to directly compare corresponding topics.²⁶

Fig. 2.3 shows the JS divergence scores for aligned topic pairs obtained using our procedure.²⁷ We choose the 44 most aligned topics as constituting a meaningful-comparison set (based on visual inspection across our three datasets). We consider

²⁶Lower JS divergence implies higher similarity, so our particular implementation aims to *minimize* the total weight. The advantage of using JS divergence is its symmetric and finite nature.

²⁷Contents of Fig. 2.3, including the description, are directly taken from Hoyle et al. (2020).

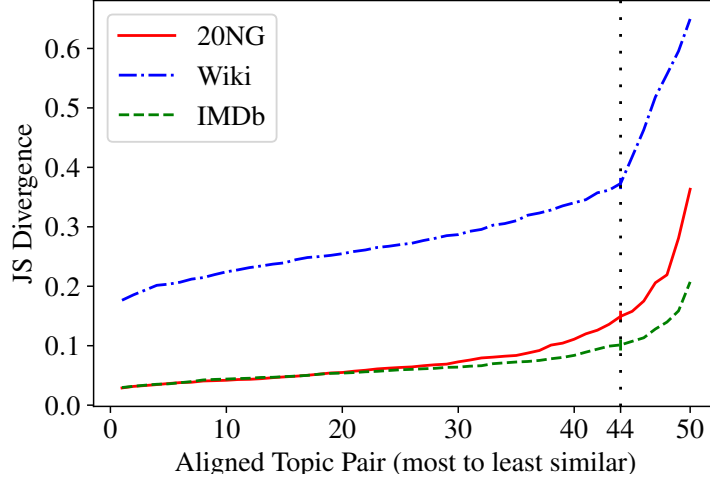


Figure 2.3: Jensen-Shannon divergence for aligned topic pairs in the SCHOLAR and SCHOLAR + BAT models.

the remaining topics as not constituting a meaningful conceptual relationship and choose the same threshold for all three datasets for simplicity.

In order to compare topics discovered by the NTM (SCHOLAR) that serves as the base model with corresponding (aligned) topics discovered by our model that is augmented with BAT, we use the NPMI values of the corresponding topics as well as examine them qualitatively. We observe that the model augmented with BAT has the topic with the higher NPMI value more often across all three datasets (Fig. 2.4) when looking at conceptually matched topic pairs.²⁸ This suggests that BAT, in addition to providing improvements in the aggregate, also offers a specific advantage of generating topics in the same space as that of an existing model (used as its base) and tends to improve the coherence of individual topics in most cases. This analysis establishes the modular value of our method.

We qualitatively discuss one example from each dataset in Table 2.5.²⁹ These

²⁸Contents of Fig. 2.4, including the description, are directly taken from Hoyle et al. (2020).

²⁹Contents of Table 2.5, including the description, are directly taken from Hoyle et al. (2020).

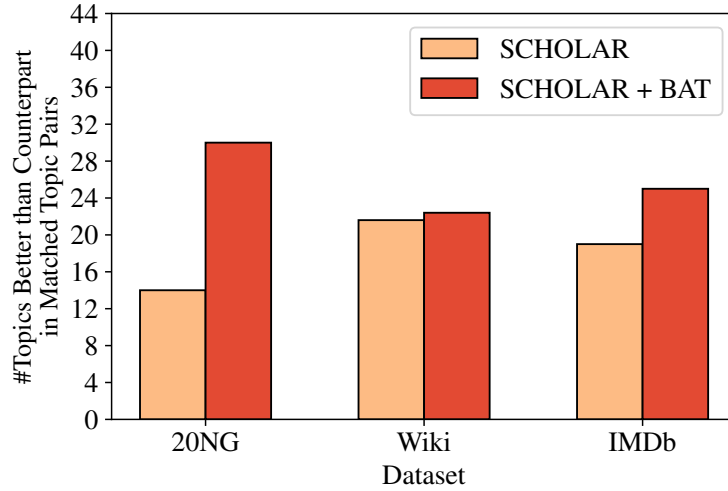


Figure 2.4: Number of matched topic pairs where SCHOLAR + BAT improves coherence, compared with the number of matched pairs where the baseline improves coherence.

examples are selected for illustration from a single run of the baseline SCHOLAR and SCHOLAR + BAT models (selected randomly) for $K = 50$. In these limited cases, the differences in NPMI values do seem to correspond to subjective differences in topic quality, according to our interpretation. Appendix A.6 provides 15 aligned topic pairs for each dataset in order to help readers form their own impressions. These pairs are selected randomly after stratification based on alignment quality scores in order to create a fair sample to review.

2.1.5 Main Takeaway

We contribute an effective approach of combining two unsupervised methods in large language models and topic models via knowledge distillation. To the best of our knowledge, our work constitutes the first instance of using a “black-box” neural network teacher to guide a probabilistic graphical model as a student. This helps us combine the expressivity and interpretability of probabilistic topic models

		NPMI	Topic
20ng	SCHOLAR	0.454	nhl hockey player coach ice playoff team league stanley european
	SCHOLAR + BAT	0.523	nhl hockey player team coach playoff cup wings stanley leafs
Wiki	SCHOLAR	0.547	jtwc jma typhoon monsoon luzon geophysical pagasa guam cyclone southwestward
	SCHOLAR + BAT	0.621	jtwc jma typhoon meteorological intensification monsoon dissipating shear outflow trough
IMDb	SCHOLAR	0.197	adaptation version novel bbc versions jane kenneth handsome adaptations faithful
	SCHOLAR + BAT	0.218	adaptation novel book read books faithful bbc version versions novels

Table 2.5: Selected examples of SCHOLAR + BAT improving on topics from SCHOLAR. We observe that the improved 20ng topic is more cleanly focused on the NHL (removing *european*, adding the Toronto Maple *Leafs*, evoking the Stanley *Cup* rather than the more generic *ice*); the improved wiki topic about typhoons is more clearly concentrated on meteorological terms, rather than interspersing specific locations of typhoons (*luzon*, *guam*); and the improved IMDb topic more cleanly reflects what we would characterize as “video adaptations” by bringing in terms about that subject (*book*, *books*, *novels*, *read*) in place of predominant words relating to particular adaptations. Randomly selected examples can be found in Appendix A.6.

with the more precise, linguistic context-driven, knowledge contained in pretrained transformers.

Our modular method can augment any neural topic model (NTM) to improve topic quality (as measured using NPMI in the aggregate as well as for individual topics). We demonstrate this using two NTMs that rely on highly distinct architectures — the variational autoencoder and the Wasserstein autoencoder — demonstrating state-of-the-art topic coherence across three different datasets that span different domains and properties.

Reporting aggregate coherence scores for model comparison is, to the best of our knowledge, a distinct default in much of neural topic model development literature. However, taking advantage of the modularity of our approach, we are able to contribute a procedure that can provide comparisons at the level of individual topics, which shows that our method improves the automatically measured coherence of most individual topics.

Our full code, including data processing, is available at: <https://github.com/hoho/kd-topic-models>.

2.1.6 Related Work

Topic models that incorporate embeddings. A major motivation behind our usage of knowledge distillation is to supply a topic model with information about relationships between different terms that may not be directly supported by the observed input documents alone. Previous topic modeling works have posited different approaches in order to integrate such external word-relationship information, such as the use of word senses (Ferrugento et al., 2016) and pretrained word embeddings (Hu and Tsujii, 2016; Yang et al., 2017; Xun et al., 2017; Ding et al., 2018). Dieng et al. (2020) provide an extensive overview of attempts to marry topic models and external word embeddings.

We also note that after our work, Bianchi et al. (2021) used BERT-based embeddings as additional input to a neural-network-based encoder. A comparison between our approach and their topic model, on criteria relevant to the real-world use case of qualitative content analysis, can be found in Hoyle et al. (2022).

Our approach utilizes both general knowledge acquired from the pretraining of the transformer model on a large external corpus that cuts across domains *and* the information specific to the corpus being modeled, by way of fine-tuning. Another advantage of our method is the modular addition that can be made with just a few lines of code (Appendix A.5) to any underlying topic model instead of constructing a specialized design as is characteristic of the above-cited methods.

Knowledge distillation. The original focus of this technique in Hinton et al. (2015) was single-label image classification, but KD has also been extended to a setting with multiple labels (Liu et al., 2018b). We refer the reader to Gou et al. (2021) for an informative survey.

Especially in NLP, application of KD has generally been limited to supervised settings (Kim and Rush, 2016; Huang et al., 2018; Yang et al., 2020b), though some unsupervised tasks have also been shown to take advantage of this technique — usually with an unsupervised teacher guiding a *supervised* student (Hu et al., 2020a; Sun et al., 2020). We note that the term distillation sometimes gets used in multiple senses, and can refer to joint learning (which is distinct from our method that follows Hinton et al. (2015)). For example, topic models have been trained jointly with word embeddings by Xu et al. (2018) where the procedure is also referred to as distillation.

Pretrained models like BERT have been shown to be good teacher models in a variety of tasks including sentiment classification and automated paraphrasing Tang et al. (2019a,b). Finally, we recall that one of our core motivations for using knowledge distillation is to retain the increased interpretability of topic models and improve the coherence of discovered topics; Liu et al. (2018a) also deploy knowledge distillation with a focus on interpretability.

Topic diversity. Several topic modeling works (Dieng et al., 2020; Burkhardt and Kramer, 2019; Nan et al., 2019) have proposed other metrics in addition to NPMI (assumed to measure coherence) that quantify the diversity or uniqueness of topics (based on top terms in topics). Not only has no one implementation of a diversity or uniqueness metric gained consensus in the literature, but these metrics also tend to suffer from a key theoretical drawback — cases where two topics have the same set of top n words (seen as essentially identical), and cases where a particular topic’s top n words are repeated individually across multiple other topics (seen as a topic with more diffuse similarity to those other topics), are treated as essentially the same. Issues related to topic diversity are discussed in more depth in Appendix A.3.1.

2.2 Automated Coherence-based Evaluation of Topic Models is Broken

Evaluating unsupervised methods without a ‘ground truth’ is challenging. Automated proxies aiming to capture broad usefulness to human users without being grounded in a clear use case can make it easy to lose sight of the goal of proposing a new method, except doing ‘better’ than previously proposed methods.

In Section 2.1.2, we presented the primary way topic models are compared when new neural topic models (NTMs) are introduced in recent methods literature — by relying on one particular class of automated metrics: coherence, often operationalized by using NPMI as used in our work above. Works proposing a NTM in relevant computer science literature compare their method with some version of LDA as a popular non-neural baseline, and based on coherence scores (and, sometimes, other automated metrics), show their proposed method as the one that finds ‘better’ or more coherent topics in the aggregate.³⁰ Then, much like our own work above, the metric is also used to automatically compare the proposed NTM with other previously proposed NTMs as baselines. By conducting a comprehensive meta-analysis (detailed below in Section 2.2.1), we find this automated evaluation paradigm using automated coherence metrics is indeed what the topic modeling development literature has coalesced around. We also find that systematic human evaluation is not done in *any* of the works introducing new NTM methods that we surveyed.

A general reason for the development and adoption of a general proxy for human judgment is to facilitate fast development and comparison without the need

³⁰Outside of the core topic modeling method development literature, human evaluations have been used to develop new metrics and improve understanding of existing model behavior (Bhatia et al., 2017; Morstatter and Liu, 2018; Lund et al., 2019; Alokaili et al., 2019, *inter alia*).

for human evaluation (which runs counter to the real-world use of topic models that often needs to be grounded in particular user applications; see [Doogan and Buntine \(2021\)](#)). In Section 2.1.2, we provide background on the development of automated evaluation metrics for topic modeling, which sheds light on why this particular choice of automated coherence measures such as NPMI was adopted. The primary reason is previous work finding a significant correlation for coherence metrics like NPMI with human judgment of topic coherence and quality ([Newman et al., 2010](#); [Aletras and Stevenson, 2013](#); [Lau et al., 2014](#)). However, topic modeling choices in these previously conducted human validation studies never used NTMs. In part, this is because these studies were conducted before NTMs were proposed and seen as viable. Instead, they relied on traditional probabilistic methods such as LDA and its variants, and, in fact, often carried out topic modeling comparisons by considering instantiations of the same topic model with different numbers of topics as the different topic modeling choices being compared.

Therefore, works comparing topic modeling methods to claim ‘better’ topic modeling based on automated coherence metrics, where these methods include NTMs, have not been directly backed by human evaluation. There is a crucial operational assumption that coherence metrics are a good or accurate proxy of human judgment even for newer architectures and modeling techniques when the metrics were only ever validated for classical methods.³¹ This suggests a *validation gap*: we do not know if the dominant evaluation paradigm based on automated metrics is still consistent with human judgment!

An additional reason to be concerned about this is the fact that in many NTM works, including our own above, the reported values on these metrics are well

³¹Here, we use “classical” to mean generative models defined by a chain of conjugate exponential family distributions optimized by Gibbs sampling or variational inference.

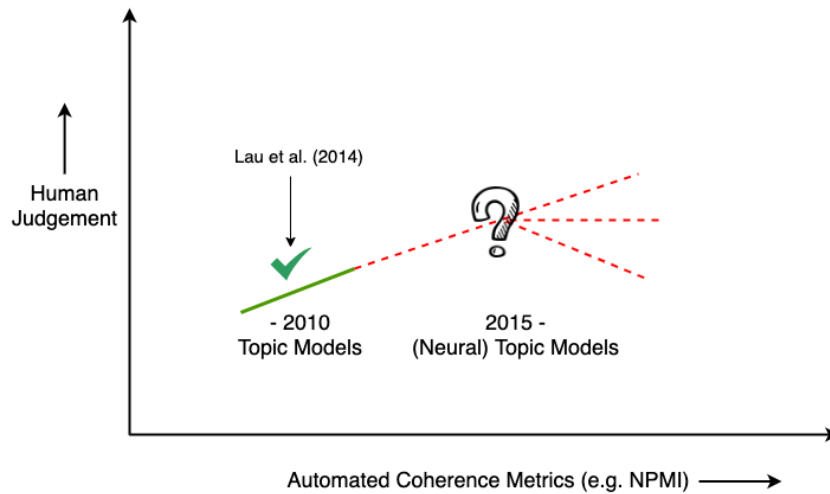


Figure 2.5: An informal illustration of the fact that the reported ranges of coherence (such as NPMI) results is well outside the range of variable comparison in [Lau et al. \(2014\)](#), which used non-neural topic modeling techniques (proposed before 2010) in order to validate automated coherence metrics with human judgments—we do not know if the same relationship that was found would hold in this new range of values (achieved using contemporary neural methods proposed after 2015).

outside the range of values used to establish the relationship between automated coherence scores and human evaluation scores in [Lau et al. \(2014\)](#) (see Fig. 2.5). This implies the presence of another *non-validated* assumption that ever-increasing automated coherence scores will continue to correspond to higher human evaluation scores.³² Further, the type of topics given by classical and neural models can be, at least in our subjective assessment, quite different (see Table 2.6).³³ This motivates skepticism that proxies validated on classical methods will continue to be a good stand-in for human judgment with newer neural methods.

By using one automated proxy as a standard in automated topic modeling evaluation, we see that the goal of most neural topic modeling literature has become improving this one particular score presumed to be a proxy for human judgment.

³²Fig. 2.5 is taken directly from the talk given at NeurIPS 2021 when presenting the work published in [Hoyle et al. \(2021\)](#).

³³Contents of Table 2.6, including the description, are directly taken from [Hoyle et al. \(2021\)](#).

	Classical			Neural		
	station	album	tropical	tropical	spore	manhattan_project
	line	band	storm	landfall	basidia	los_alamos_laboratory
	bridge	music	hurricane	cyclone	spores	robert_oppenheimer
	railway	song	cyclone	utc	mycologist	enrico_fermi
	trains	released	depression	weakening	hyphae	physicist
NPMI	0.274	0.285	0.394	0.446	0.456	0.470

Table 2.6: The first three columns are the highest-NPMI topics for a classical topic model (LDA estimated via Gibbs sampling using Mallet, [McCallum, 2002](#); [Griffiths and Steyvers, 2004](#)). The next three are counterparts from a neural model (our D-VAE reimplementation, [Burkhardt and Kramer, 2019](#)). Models are trained on Wikitext ([Merity et al., 2017](#)) with fifty topics, and NPMI is estimated over the top five words in each topic using a 4.6M-document reference Wikipedia corpus. The mean top-five NPMI over all topics is 0.156 for the classical and 0.256 for the neural model.

Such a myopic focus can actually lead to lower quality results ([Stiennon et al., 2020](#)) and raise the possibility captured in Goodhart’s law —“when a measure becomes a target, it ceases to be a good measure” ([Strathern, 1997](#)) — a possibility we investigate in this section.³⁴

Apart from the validation gap, we also find many experimental consistencies in NTM literature. This includes failures to specify enough information about particular metrics being used (such as the choice of reference corpus for retrieving word co-occurrence counts), the particular implementation of the baseline being used, hyperparameters, etc. — details that are necessary for replication. In addition, different data preprocessing mechanisms are used both across different works and sometimes within the same work. Perhaps even more alarmingly, we find that in the majority of cases, the hyperparameters for baseline topic models used in the study proposing a new neural method do not get tuned on the datasets used in the work, which makes the fairness of any ensuring comparison and generalizability as

³⁴[Hoyle et al. \(2022\)](#) highlight other possible misleading takeaways enabled by targeting a particular proxy, by considering different evaluation criteria.

well as reliability of any claims or findings questionable (Section 2.2.1). This can be seen as a *standardization gap*.

We ensure consistency in preprocessing in our experiments above (Section 2.1.3) and tune our baseline methods (not just our proposed method) — we highly recommend these practices. Publicly available preprocessed datasets can help ensure standardization, and we make two popular processed English language datasets available in this portion of our work to help benchmarking efforts. We recommend researchers use recent preprocessing pipelines such as <https://github.com/ahoho/topics> for all their datasets, and ensure the same data and other contextual settings are being used for each topic model being compared.³⁵ In order to ensure such consistency across models in terms of both data and hyperparameter optimization, recent work by Terragni et al. (2021) has introduced a comprehensive end-to-end framework with the direct purpose of facilitating topic modeling comparison, which can help address the standardization gap in the literature.

While our meta-analysis below sheds light on the standardization gap, we will largely focus on the validation gap in this section by conducting human evaluation comparing different topic modeling outputs and measuring the correlation between human judgment and automated coherence-based evaluation of topic models.

Roadmap for this section. After summarizing and discussing the findings of our meta-analysis of neural topic modeling literature (Section 2.2.1), we discuss the previously proposed human measures of topic coherence that we adopt in this work (Section 2.2.2), AND explain the dataset, processing, and topic modeling method choices for this particular part of our work (Section 2.2.3). We then elucidate our human evaluation process (Section 2.2.4). Finally, we present the experimental

³⁵Credit: Alexander Hoyle put together this pipeline and released it as part of this joint work.

Evaluation	Count		Experimentation	Count	
Number of human evaluations	0	(0%)	<i>Preprocessing</i>		
<i>Automated Coherence</i>			Inconsistent over datasets	12	(30%)
<i>Metric</i>			Ambiguous preprocessing	9	(23%)
NPMI	26	(72%)	<i>Model comparisons</i>		
Other	22	(61%)	All models tuned	5	(13%)
Explicit implementation	22	(61%)	Unclear h.param search	16	(40%)
Explicit ref. corpus	10	(28%)	Unclear LDA baseline, if used	7	(24%)
Perplexity w/o coherence	3	(8%)	Recent baseline (w/in 2 yrs)	31	(78%)
			Multiple runs / sig. testing	11	(28%)

Table 2.7: Meta-analysis of forty neural topic modeling papers (denominator may change, as not all conditions are applicable). No recent neural topic modeling papers use human evaluations of coherence, and the metrics and models are difficult to replicate.

findings of our investigation of the potential validation gap in detail along with a discussion of the implications (Section 2.2.5), before concluding this portion of the chapter in Section 2.2.6.³⁶

2.2.1 Meta-analysis of Neural Topic Modeling

In order to gain an understanding of the evaluation paradigm dominant in topic modeling development, we conduct a survey of the neural topic modeling (NTM)

³⁶The description of the work in Section 2.2 is adapted from joint work done in: **Pranav Goel***, Alexander Hoyle*, Denis Peskov*, Andrew Hian-Cheong*, Jordan Boyd-Graber, and Philip Resnik. “Is Automated Topic Model Evaluation Broken?: The Incoherence of Coherence.” *Advances in Neural Information Processing Systems* (2021). (* indicates equal contribution)
Goel’s primary contributions to that work include:

- Jointly with his colleague (Hoyle), from their direct experiences in the previous portion of work detailed in Section 2.1, fundamental involvement in understanding the possible gaps in automated evaluation, which motivated and directly led to this part of the work.
- Helping implement the comprehensive meta-analysis of neural topic modeling literature and summarizing the results.
- Jointly with his colleague (Hoyle), setting up and implementing the experimental baselines.
- Understanding and explicating prior literature on both automated evaluation measures and the human evaluation measures we adopted in this work, and more broadly, contextualizing and situating our efforts in the literature.
- Identifying relevant examples and creating some of the figures used to convey our findings.

literature. First, we use an existing, comprehensive survey of NTMs ([Zhao et al., 2021a](#)) and extract all their references. Next, from that superset, we select (a) topic modeling papers that (b) mention topic interpretability (and therefore demonstrate an interest in evaluating a model based on how interpretable the topics are), and (c) conduct a comparison of their proposed models’ topic quality with an existing baseline model. This process results in forty models that all claim superior or state-of-the-art topic coherence.

Finally, we examine data processing steps, hyperparameter tuning practices, their procedure for selecting baseline(s), and the specific calculations and metrics used for obtaining topic coherence values. The results of our survey are summarized in Table 2.7.³⁷ We provide all the papers used in this survey in Appendix A.7.

We find variation in all of the aforementioned areas. Within the same paper, preprocessing is often (30%) inconsistent across datasets, which is significant since preprocessing can meaningfully affect the model quality and results provided by automated metrics. When preprocessing *is* consistent, details required to replicate the pipeline are not always provided. Such issues suggest that the computation of automated metrics, even for the same baseline models and source datasets, varies across different papers.

The problem is compounded by the practice of authors training their models on datasets that are different from the ones that were used to establish the relationships between human annotations and automated metrics, which provided the credibility needed for the adoption of those metrics. [Doogan and Buntine \(2021\)](#) find that the association of the same metrics with interpretability may not generalize to new domains. Per our analysis, 40% of papers do not clearly specify their model tuning procedure and often not even the metric that was used for model selection; this

³⁷Contents of Table 2.7, including the description, are directly taken from [Hoyle et al. \(2021\)](#).

mirrors the findings in [Dodge et al. \(2019\)](#).

The calculation of automated coherence metrics is also riddled with similar problems. As we detail in Section 2.1.2, calculating NPMI values involves several specific steps and decisions such as the reference corpus used to estimate joint word probabilities, size of the context window used to obtain co-occurrence statistics, and the number of top words in a topic. A remarkable 75% of the surveyed papers do not explicate their reference corpus. Even if we grant the benefit of the doubt and assume that the input corpus is being used (13 cases), it is not clear if the authors used a held-out set or the training documents themselves for their calculations, or a different strategy altogether.

Considering the 61% of cases where the implementation of the coherence metric is specified through an explicit formula or a pointer to a code repository, some of the aforementioned factors may still not be completely clear. This is the case, for example, in six of the papers where a reference to [Lau et al. \(2014\)](#) is made along with that work’s code, but that does not make things completely unambiguous since [Lau et al. \(2014\)](#) recommend the usage of a reference corpus that comes from the same source as the specific training data being uses whereas the repository itself defaults to Wikipedia.³⁸ In other cases, authors use very particular implementations that either create room for errors or deviate from human experimentation settings in [Lau et al. \(2014\)](#). For example, several papers use a document-wide context window for calculating NPMI scores – a choice that is distinct from the implementation choice made in works that strive to show a correlation with human judgments.

Finally, we note that beyond the consistency-related issues, virtually all claims made in this work regarding topic models that offer higher coherence rest on the automated metric’s validity, established in [Lau et al. \(2014\)](#), generalizing to NTMs.

³⁸https://github.com/jhlau/topic_interpretability

This constitutes the critical validation gap discussed as the core motivation for this work at the start of this section.

2.2.2 Human Measures of Topic Coherence

Below we review two metrics or methods for quantifying human judgments of topic quality: direct ratings and intrusion. These metrics were used in [Lau et al. \(2014\)](#) to obtain human scores, and particularly when scores were aggregated across all topics discovered by a model, they were found to be correlated with some automated metrics. As noted earlier, this particular validation effort directly led to the widespread adoption of pairwise NPMI as *the* automated coherence evaluation metric of choice in order to compare topic models, but crucially, no NTMs were used in that study.

Rating. Raters see a topic and then give the topic a quality score, conventionally on a three-point ordinal scale ([Newman et al., 2010](#); [Mimno et al., 2011](#); [Aletras and Stevenson, 2013](#), *inter alia*).³⁹

Intrusion. [Chang et al. \(2009\)](#) devised *word intrusion* as a behavioral assessment of topic coherence. The core idea is that when the top words in a topic identify a coherent latent category, it is easier to identify words that do not belong to that category. Operationally, each topic is represented as its top words plus one “intruder” word, which has a low probability of belonging to that topic but a high probability of belonging to a different topic. A good topic is one where human annotators reliably detect the “intruder” word.⁴⁰

³⁹We also present correlation findings using a five-point ordinal scale in Appendix [A.13](#).

⁴⁰Descriptions of the *rating* and *intrusion* measures are taken directly from [Hoyle et al. \(2021\)](#).

	WIKI	NYT
Domain	Encyclopedia	News
<i>Number of Docs.</i>		
Training	28.5k	273.1k
Reference	4.62M	1.82M
Mean Tokens / Doc.	1291	281
Vocab. Size	39.7k	34.6k

Table 2.8: Corpus statistics. Datasets vary in domain, average document length, and total number of documents. WIKI is from [Merity et al. \(2017\)](#) and NYT is from [Sandhaus \(2008\)](#).

2.2.3 Experimental Detail

We discuss our choices of dataset, processing, and models used to set up our human evaluation experiments (and the eventual comparison with automated metrics at the heart of this study).

We process two commonly-used datasets using a standard pipeline. We then use a computationally fair hyperparameter search paradigm in order to estimate topic models on each dataset. This helps us address the standardization gap in topic modeling literature discussed above.⁴¹

Datasets and Preprocessing

Per the experiments in [Chang et al. \(2009\)](#), which is the first instance of topic modeling work that sought to validate automatic metrics of modeling quality using human judgments, we utilize English articles from Wikipedia (WIKI) and the New York Times (NYT). We provide corpus statistics in Table 2.8.⁴²

For WIKI, we use Wikitext-103 ([Merity et al., 2017](#)), and for NYT, we subsample roughly 15% of documents from LDC2008T19 ([Sandhaus, 2008](#)) (which makes it an

⁴¹We also refer readers to similar standardization efforts in [Terragni et al. \(2021\)](#).

⁴²Contents of Table 2.8, including the description, are directly taken from [Hoyle et al. \(2021\)](#).

order of magnitude larger than WIKI). Reference counts are computed using a 4.6 million document Wikipedia dump (from September 2017) and the full 1.8 million document LDC2008T19 New York Times set – both these corpora are processed in an identical manner to the training data.

We use SpaCy ([Honnibal et al., 2020](#)) for tokenization and entity identification and create new tokens for detected entities of the form `New_York_City`, per [Krasnashchok and Jouili \(2018\)](#). We do not lemmatize the text following observations in [Schofield and Mimno \(2016\)](#) demonstrating that lemmatization and word-stemming can hurt English topic interpretability.

In order to maintain a roughly equal vocabulary size across datasets, we utilize a power-law relationship for corpus size (c.f. [Zipf, 1949](#)) and rule out tokens occurring in fewer than a given number of documents.⁴³ We use both a standard stopwords list and a corpus-specific stopwords list consisting of tokens that appear in more than 90% of documents. Complete preprocessing details are provided in Appendix [A.8](#).

Models

Much of the NTM literature compares a newly proposed model along with some baseline NTM models and a classical or non-neural traditional model, usually LDA. We, therefore, validate the automated coherence metric using one classical model and two newer NTMs:

Gibbs-LDA (G-LDA). We use the widely-adopted Mallet ([McCallum, 2002](#)) implementation for Gibbs-sampling-based LDA ([Griffiths and Steyvers, 2004](#)) as a strong classical baseline.

⁴³We target vocabulary sizes that approximate the number of words known by an adult English-speaker ([Brysbaert et al., 2016](#)), which translates to roughly 40k for WIKI and 35k for NYT.

Dirichlet-VAE (D-VAE). We reimplement Dirichlet-VAE (Burkhardt and Kramer, 2019), which is demonstrated to be state-of-the-art NTM in terms of automated coherence scores.⁴⁴ Dirichlet-VAE provides a meaningful improvement on one of the first successful NTMs — (PRODLDA, Srivastava and Sutton, 2017) — and showcases competitive automated coherence scores against recent models. Further, the generative model utilized retains a broad similarity to LDA, especially on account of using a Dirichlet prior. The primary advantage of this model is that it does not limit the estimated topic-word distributions to the simplex.

ETM. Many NTMs try to take advantage of their flexibility in order to incorporate external word representations, motivated by the possibility of using large-scale, general language knowledge for improving topic quality. Indeed, this was highlighted by our own proposed method (Section 2.1.1). The Embedded Topic Model (Dieng et al., 2020) is a popular NTM that uses word embeddings in its generative model.⁴⁵

Implementation Details. Following the extensive auditing conducted by Dodge et al. (2019), we use a fixed computational budget for each model and use a random set of 164 hyperparameter settings across datasets and model type.⁴⁶ Each model is trained for a variable number of steps (a hyperparameter), and the automated coherence score for the model in question is calculated using topics produced at the last step.

⁴⁴For simplicity, we use pathwise gradients for the Dirichlet (Jankowiak and Obermeyer, 2018a) instead of the rejection sampling variational inference used in the authors’ primary variant in Burkhardt and Kramer (2019). We replicate the NPMI and redundancy scores on the 20 newsgroups dataset: github.com/ahoho/dvae.

⁴⁵github.com/adjidieng/ETM; paper citations and GitHub stars of the repository indicate the popularity of this model.

⁴⁶We note that runtimes vary drastically across models, though our particular study is not concerned with the efficiency of the implementation, although efficiency is certainly a critical concern (see Ethayarajh and Jurafsky, 2020).

For human evaluations, we select models that maximize NPMI, which is calculated using the reference corpus (as specified above), with the number of top topic words used set at 10, and a context-window size for obtaining co-occurrence accounts also set at 10 following [Lau et al. \(2014\)](#).

Following [Dieng et al. \(2020\)](#), skip-gram embeddings are learned on the training corpus for the ETM model, although we note that experiments using external pretrained embeddings did not result in substantially different results.

As we did in proposing our BAT model above (Section 2.1.1, Section 2.1.3), we do not consider models where the output degenerates and results in a highly redundant set of topics, a known possibility for NTMs ([Burkhardt and Kramer, 2019](#)). These discarded models are either the ones where the top five words of one topic overlap with another topic or the ones with a topic uniqueness score ([Nan et al., 2019](#)) above 0.7. We provided other details for model implementation including ranges for conducting optimal hyperparameter value searches in Appendix A.9.

2.2.4 Process for Conducting Human Evaluation

The validation gap involves a lack of human evaluation with the topic modeling outputs of NTMs and comparing those judgments with the results obtained using NPMI. Addressing this gap is the goal of our human evaluation study.

We use the *ratings* and *intrusion* tasks (described above in Section 2.2.2) as human evaluations of topic quality following prior work. We recruit crowdworkers using Prolific.co (an online panel provider) and use the Qualtrics survey platform for data collection. We pay workers 2.5 USD per ratings survey and 3 USD per word intrusion survey, which is equivalent to 15 USD per hour.

Drawing meaningful conclusions from human annotations requires an adequate

Select which term is the least related to all other terms and your familiarity with the words

Terms

- ☐ painting
- ☐ paintings
- ☐ casualties
- ☐ painter
- ☐ literary
- ☐ poems

Answer Confidence

- ☐ I am familiar with most of these terms.
- ☐ I am **not** familiar with most of these terms, but I **can** answer confidently.
- ☐ I am **not** familiar with most of these terms, and so I **cannot** answer confidently.

Figure 2.6: The word intrusion task presented to crowdworkers (the ratings task is shown in Fig. 2.7).

Please rate how related the following terms are to each other and how familiar you are with the terms

concerto, balanchine, mozart, orchestra, brahms, beethoven, recital, choreographers, schubert, composers

Rating

- ☐ Very related
- ☐ Somewhat related
- ☐ Not very related

Answer Confidence

- ☐ I am familiar with most of these terms.
- ☐ I am **not** familiar with most of these terms, but I **can** answer confidently.
- ☐ I am **not** familiar with most of these terms, and so I **cannot** answer confidently.

Figure 2.7: Ratings task presented to crowdworkers.

number of participants to conduct a statistically well-powered experiment. [Card et al. \(2020\)](#) note that insufficient power for a robust examination of model differences is a problem that applies to many NLP experiments, including those relying on human evaluation. In order to ensure enough statistical power (at least $1 - \beta = 0.9$), we use a straightforward generative model of annotations and then select enough crowdworkers per task in order to obtain significance at $\alpha = 0.05$. This results in a minimum of 15 crowdworkers per topic for both tasks. We note that based on this criterion, the experiments in [Chang et al. \(2009\)](#) as well as [Lau et al. \(2014\)](#) are underpowered (both used eight annotators).

We obtain fifty topics each from all the three models detailed in Section 2.2.3 across both of our datasets. For the word intrusion task, we sample five out of the

top ten topic words along with one intruder (Fig. 2.6).⁴⁷ In the ratings task, we present the top ten words in order and obtain the Likert-scale ratings (Fig. 2.7).⁴⁸

For each of the two tasks and the two datasets, we randomly sample 40 of the 150 topics generated in total by the three topic models. For the ratings task, we additionally include sixteen synthetic topics of poor quality, which helps us calibrate the scores obtained from annotations and filter out low-quality respondents.⁴⁹

We follow Chang et al. (2009) in designing the wording of our questions, and ensured that the crowdworkers received detailed instructions along with examples before responding to items.⁵⁰ These detailed instructions are included in Appendix A.10. Since topics can sometimes be esoteric due to the input data (for e.g., see the last columns of Table 2.6), we ask crowdworkers about their familiarity with the words in each question. This question might help protect against spurious low scores for otherwise coherent topics, since real-world users of topic models are usually familiar with domain-specific terminology (further discussed in Section 2.2.6).

2.2.5 Results: Human Judgments Differ from Automated Metrics

Using our three selected topic models that provide topic estimates, we compare human judgments to automated metrics.

⁴⁷Contents of Fig. 2.6, including the description, are directly taken from Hoyle et al. (2021).

⁴⁸Contents of Fig. 2.7, including the description, are directly taken from Hoyle et al. (2021).

⁴⁹For generating *synthetic poor-quality topics*, we use random high-probability words appearing in topics that were obtained for topic models utilizing other, non-optimal, hyperparameter settings, but that have low probability among selected topics, in order to keep the synthetic topics faithful to the data while representing low quality. Eight topics each are generated from the vocabularies of NYT and WIKI.

⁵⁰Code that converts the topic model output into deployable questionnaires is provided at github.com/ahoho/topics.

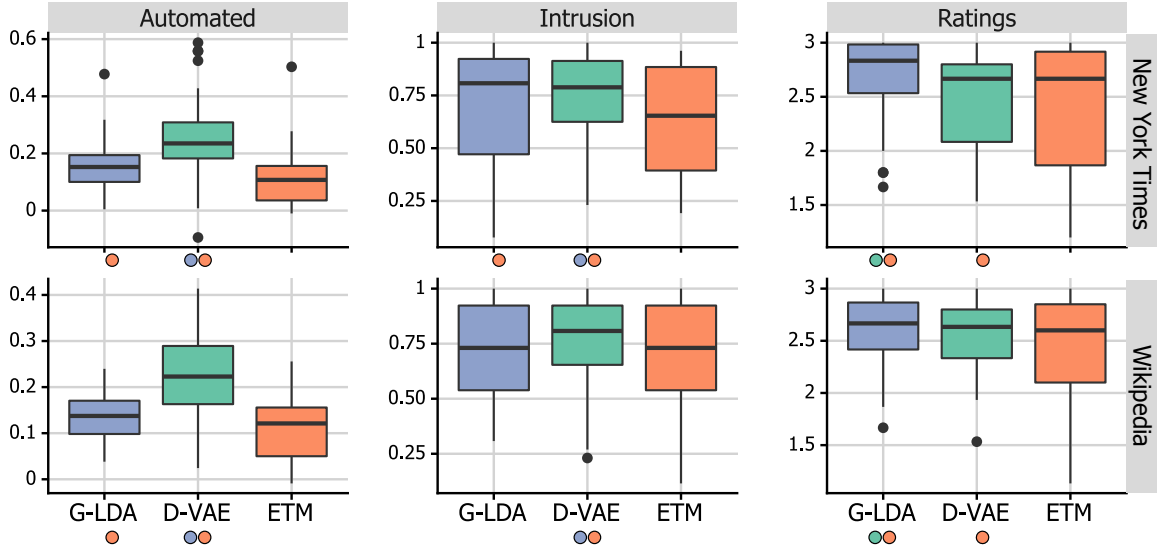


Figure 2.8: While automated evaluations (here, NPMI) suggest a clear winner between models, human evaluation is more nuanced. Human judgments exhibit greater variability over a smaller range of values. Colored circles correspond to pairwise one-tailed significance tests between model scores at $\alpha = 0.05$; for example, the rightmost orange circle at the bottom right shows that human intrusion ratings for D-VAE are significantly higher than ETM for topics derived from Wikipedia.

Human Assessment. We establish differences between each pair of topic models via pairwise significance tests using human ratings: a proportion test for the intrusion scores, a U test (Mann and Whitney, 1947) for the ratings, and a t -test for automated metrics. We use one-tailed tests for each pair in both directions. The results are shown in Fig. 2.8.⁵¹ The word *intrusion* task suggests D-VAE compares favorably with the other two models across both datasets, but G-LDA is favored when using the *ratings* human evaluation task.⁵²

We note the consistency of our human evaluation results with previous iterations of the ratings and word intrusion tasks used to compare topic models. Specifically, Mimno et al. (2011) find an average score of 2.36 for the ratings task when using

⁵¹Contents of Fig. 2.8, including the description, are directly taken from Hoyle et al. (2021).

⁵²These discrepancies among human tasks suggest that using standard coherence metrics alone may be *insufficient* for automatically selecting a suitable ‘best’ topic model among many options, which is also suggested in Doogan and Buntine (2021).

Ref. Corpus → Train Corpus ↓		NPMI (10-token window)				C_v (110-token window)			
		NYT	WIKI	Train	Val	NYT	WIKI	Train	Val
Intrusion	NYT	0.27	<u>0.43</u>	0.27	0.24	0.34	<u>0.45</u>	0.35	0.34
	WIKI	<u>0.34</u>	<u>0.36</u>	<u>0.39</u>	0.17	<u>0.32</u>	<u>0.34</u>	<u>0.34</u>	0.20
	Concatenated	0.29	<u>0.40</u>	0.32	0.17	0.32	<u>0.40</u>	<u>0.35</u>	0.24
Rating	NYT	0.37	<u>0.48</u>	0.37	0.39	<u>0.41</u>	<u>0.46</u>	<u>0.44</u>	<u>0.45</u>
	WIKI	0.34	<u>0.41</u>	<u>0.44</u>	0.28	0.32	<u>0.40</u>	<u>0.40</u>	<u>0.34</u>
	Concatenated	0.37	<u>0.44</u>	<u>0.41</u>	0.35	0.38	<u>0.42</u>	<u>0.42</u>	<u>0.42</u>

Table 2.9: Spearman correlation coefficients between mean human scores and automated metrics. Underlined values have overlapping bootstrapped 95% confidence intervals with that of the **largest** value in each row. “Concatenated” refers to correlations computed on a concatenation of values for the NYT and WIKI items. “Val” is a small held-out set of 15% of the training corpus. Using the more data-appropriate logistic and ordered probit regressions for word intrusion and ratings data leads to different conclusions about relative metric strength (appendix table A.16). CIs are estimated using 1,000 samples.

a dataset of medical paper abstracts, and our score averages range from 2.5 to 2.8 across all our variations (Fig. 2.8).⁵³ For the word intrusion task, the mean scores we obtain range from 0.7 to 0.8, comparable to the intrusion score of ~ 0.8 obtained for the LDA model in Chang et al. (2009). The median time taken on the tasks was 8 – 9 minutes.

Following Aletras and Stevenson (2013), we use the Spearman correlation between each respondent’s score (per topic) and the average of all the other respondent scores in order to calculate the inter-annotator agreement. We achieve an agreement score of 0.75, which is comparable to Aletras and Stevenson’s reported value of 0.7 on the NYT corpus. We control for insincere submissions by including synthetic poor-quality topics (footnote 49), that were correctly identified by annotators, and we also monitor the total time taken to complete the survey.

⁵³Newman et al. (2010) and Lau et al. (2014) do not report an average.

Automated Metrics. While NPMI finds D-VAE as the clear ‘winner’ among the three topic models (and G-LDA as a clear second; Fig. 2.8), human judgments convey a very different story and do not find such clear winners.

We estimate the Spearman correlation between the two sets of values across each task and dataset for the metric variants (Table 2.9) in order to help understand the relationship between automated metrics and human ratings.⁵⁴ Prior studies rely on mean human ratings across all topics but that can obscure the variance of human ratings, which can result in overconfident estimates. We correct for this potential shortcoming by instead resampling ratings (with replacement) to construct 95% confidence intervals. We resample as many times as the number of annotators per task (Table 2.9).

We use the standard 10-word context window for calculating NPMI and calculate C_v with the recommended 110-word window as per the metric’s introduction in Röder et al. (2015).⁵⁵

Interestingly, our findings show that the Wikipedia corpus is best correlated with human judgments, even when the models are trained on NYT; this is in contrast to the findings of Lau et al. (2014), which finds that within-domain data has the highest correlations.

While all the correlation coefficients we obtain are statistically significant, we *urge* the need to move beyond just the strength of the correlation as a basis for justifying their use for model selection, which, per our survey and meta-analysis, is the standard in NTM literature (Section 2.2.1). Even if the direction of change matches (as measured by the correlation), it can be difficult to ascertain exactly *when*

⁵⁴Contents of Table 2.9, including the description, are directly taken from Hoyle et al. (2021).

⁵⁵We use the Gensim library in Python (Řehůřek and Sojka, 2010) to calculate coherence. Reference corpora are processed in the same manner as the training data. Other metrics, like C_{UCI} (Newman et al., 2010) and C_{UMASS} (Mimno et al., 2011), show low correlations.

an increase in the average automated coherence score for a model constitutes a significant improvement in the corresponding human scores, owing to the inherent uncertainty in human judgments. To take a step towards a remedy for this uncertainty and its implications, we provide results using one possible better model of human scores — GLMs — in Appendix A.11 and Table A.16.

As we note above, our findings indicate that automated metrics exaggerate model differences compared to human judgments (Fig. 2.8). In order to further assess this phenomenon and gain an understanding of the (in-)applicability of automated metrics for model comparison, we estimate the number of times an automated metric incorrectly finds one model to be better than another. This corresponds to an estimate of the false discovery rate for each model, obtained via bootstrapping. For each dataset, we draw random samples of two independent sets of $K = 50$ topics (without replacement) from the 150 topics across the three topic models. We retain the corresponding automated and human scores. Considering the two sampled sets as outputs from two different models, we repeat the computation of pairwise significance tests between each set for $N = 1000$ iterations. Pairwise tests are conducted for both the K automated metrics and $K \times M$ human scores (using proportions z -test for the intrusion scores and t -tests for all other values). We find that for even the best-performing automated metrics, significant differences are reported despite equivalent human scores roughly one-fifth of the time (Table 2.10).⁵⁶ This constitutes a notable discrepancy between automated and human evaluation of topic modeling (and as criteria for model selection).

Discussing the Discrepancy between the Results of Human and Automated Measures. We find that an individual topic’s NPMI or C_v score has a significantly

⁵⁶Contents of Table 2.10, including the description, are directly taken from Hoyle et al. (2021).

		NPMI (10-token window)			C_v (110-token window)		
	Ref. Corpus → Train Corpus ↓	NYT	WIKI	Train	NYT	WIKI	Train
Intrusion	NYT	46 / 53	34 / 48	48 / 50	35 / 38	30 / 29	34 / 35
	WIKI	44 / 76	33 / 78	33 / 75	45 / 48	38 / 49	37 / 45
	Concatenated	42 / 67	40 / 66	41 / 64	36 / 46	31 / 44	30 / 45
Rating	NYT	45 / 50	45 / 51	41 / 47	27 / 29	26 / 26	21 / 26
	WIKI	40 / 73	31 / 73	33 / 71	38 / 40	31 / 40	28 / 34
	Concatenated	39 / 66	36 / 66	37 / 62	31 / 38	28 / 38	19 / 36

Table 2.10: False discovery rate (1 – precision, lower is better) and false omission rate of significant model differences when using automated metrics; automated metrics often overstate meaningful model differences. **Bolded** values are those with the lowest geometric mean of FDR and FOR. We sample two independent sets of 50 topics along with their human scores and automated metrics; these sets act as the outputs of two “models”. We then compute significance tests between sets (per Fig. 2.8) on both the automated scores and human scores. A false positive occurs when one set has significantly larger automated scores despite no meaningful difference in actual human scores. Estimates are over 1,000 samples.

Data	Model	Topic	NPMI	Rat.	Int.
NYT	D-VAE	inc 6mo earns otc rev qtr 9mo nyse outst dec	0.56	1.60	0.77
WIKI	D-VAE	waterline conning turrets boilers amidships aft knots armament guns mounts	0.33	1.93	0.65
NYT	G-LDA	bedroom room bath taxes year market listed kitchen broker weeks	0.30	2.00	0.23
NYT	D-VAE	condolences mourns mourn board_of_directors heartfelt deepest esteemed	0.38	2.60	0.23
NYT	D-VAE	shareholders earnings federated mci shares takeover new_york_stock_exchange	0.18	3.00	0.81
WIKI	D-VAE	continental_army expedition militia frigate musket frigates muskets skirmish	0.11	3.00	0.69
NYT	D-VAE	medicaid medicare hospitals welfare uninsured patients	0.13	2.80	0.96
NYT	G-LDA	city mayor state new_york new_york.city officials county yesterday governor	0.09	2.53	1.00

Table 2.11: Topics with the largest human–NPMI discrepancies; top half are topics where NPMI is high and human preferences are low, bottom half is the reverse. NPMI favors esoteric and corpus-specific topics. NPMI is calculated with a 10-token sliding window over the in-domain reference corpus, **Rat.** is the average 3-point rating for a topic, and **Int.** refers to the percentage of annotators who identify the intruder word.

negative correlation with the share of annotators who report familiarity with topic words (Pearson’s $\rho = -0.29$). In the aggregate, we also find that the D-VAE generates topics with the fewest familiar words for human annotators even though it scores the highest on automated measures out of all three models; in fact, across both tasks, annotators are familiar with the terms of topic words presented to them over

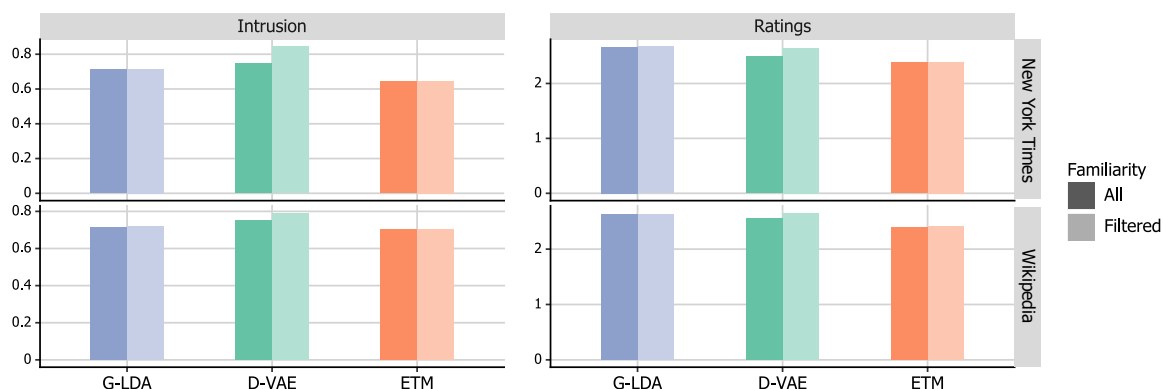


Figure 2.9: Mean human evaluation on the ratings and word intrusion tasks, after filtering out respondents who reported a lack of familiarity with the topic words. When filtering, D-VAE scores improve, highlighting its tendency to produce esoteric topics.

90% of the time for both G-LDA and ETM but only 70% of the time for D-VAE. This suggests that automated metrics favor more esoteric topics, which might constitute one reason for the observed discrepancy between humans and automated measures.

Further, this finding indicates that the D-VAE model selects narrower topics that are more specific or restricted in their scope than the other models. In fact, removing the item annotations where annotators are unfamiliar with the presented terms substantially increases the word intrusion task accuracy as well as the proportion of cases marked as “Very related” in the ratings task (Fig. 2.9).⁵⁷

A qualitative examination of topics with a high NPMI but low human ratings makes this observation clearer. In the top half of Table 2.11, the top row consists of financial terms that frequently co-occur in NYT articles while the second row consists of rare terms about boating; both can be viewed as reasonable topics for their respective corpora.⁵⁸ In the bottom half of the same table, we observe instances where words are qualitatively very related but that is not captured by the NPMI scores (which are quite low). This could be due to these related words simply not

⁵⁷Contents of Fig. 2.9, including the description, are directly taken from Hoyle et al. (2021).

⁵⁸Contents of Table 2.11, including the description, are directly taken from Hoyle et al. (2021).

frequently co-occurring within a ten-word sliding window (Eq. (2.5)).

When considering cases with familiar terms, it is also possible for some topics to be sensible in the specific context of a corpus even when the top words characterizing the topic do not share an obvious semantic relationship. For example, in the third and fourth rows in the top half of Table 2.11, the topics contain component terms that are somewhat unrelated (e.g., “taxes” and “bedroom” in the third row), but, crucially, in the context of the New York Times, they can indeed be expected and represent coherent concepts characterizing many of the articles in that publication: real estate listings and obituaries. These kinds of topics also increase the difficulty of the word intrusion task. For these two topics in particular, only 23% of crowdworkers were able to identify the intruder word for both topics.

We note that even after accounting for term familiarity by filtering out respondents not familiar with topic terms (as a proxy for domain expertise), automated metrics continue to overstate the differences between topic models (Appendix A.12). This suggests that the problems with topic model evaluation might even extend to the method of *human* evaluations and not just automated ones.⁵⁹

2.2.6 Main Implications

Our main finding in this portion of the work is that the dominantly used automated evaluation paradigm declares a ‘winning’ topic model when corresponding human evaluations do not, calling into question the validity of conducting fully automated evaluations independent of human judgments. More specifically, automated metrics like NPMI have been used to say that proposed neural methods

⁵⁹Another relevant research approach to topic modeling intends to help humans interact with topic modeling output and help refine it; see Lee et al. (2017) for an overview on how non-expert humans understand topic modeling output (and work to refine it when enabled to do so).

produce more interpretable or coherent topics (on average) as compared with classical methods such as LDA. We conduct human evaluation of LDA and two recent neural topic models, and find that in terms of two types of human judgments of topic quality, humans, on average, do not find a significant difference between the two types of models. Therefore, automated metrics seem to be systematically exaggerating differences in modeling quality.

Our results suggest that automated topic modeling evaluation needs to be reexamined.⁶⁰ This is concordant with the findings of [Doogan and Buntine \(2021\)](#), who find that “coherence measures designed for older models [...] may be incompatible with newer models.” In order to conduct this reexamination, we need to recognize that both automatic and human evaluations of a computational method are abstractions of some real-world procedure and problem. For example, the popular criterion known as precision-at-10 in information retrieval is meant to correspond to a user who is only willing to consider the top ten retrieved documents.⁶¹

Understanding this real-world context, we assert that instead of a methods-driven approach, a *needs*-driven approach should be considered for topic modeling evaluation.⁶² This is exactly what is explored in [Hoyle et al. \(2022\)](#), by grounding evaluation criteria in the needs of qualitative content analysis as the use case. After establishing qualitative content analysis as a dominant use case for topic models and identifying the criteria for ‘good’ content analysis, model stability and alignment with human-discovered categories are identified as critical measures for adopting a topic model for assisting with content analysis. The results in that work indicate

⁶⁰Note that human evaluation itself might also need reexamination (Section 2.2.5), but in this particular work, we do use existing evaluation procedures in order to be consistent with prior work.

⁶¹This example was also used in our work in [Hoyle et al. \(2021\)](#).

⁶²Practically, ‘needs’ also has a computational component: neural models, even with GPUs, usually have longer runtimes. This contrasts with the local, CPU-only environment that many practitioners tend to use. We provide additional details on runtimes for various topic models in Appendix A.9.

that neural topic models (NTMs) fare worse than the classic LDA method, which further highlights how the field of topic modeling methods development may have been misled by using NPMI as a proxy for comparing topic models.⁶³

We note that an important limitation that applies to this entire chapter is the usage of only English datasets, when topic modeling has been used to characterize texts in many languages. While *theoretically* we expect our findings to generalize beyond the English language, *empirical* generalizability across languages remains an open question.⁶⁴

⁶³We note that our proposed knowledge distillation-based BAT method (Section 2.1) also fares worse than LDA on the criteria of model stability and alignment with human-discovered categories (Hoyle et al., 2022). However, in a specific use case explored by Gao et al. (2021), our proposed BAT method, which is a neural topic model, fares better in expert evaluation of topic modeling of discussions on maternal health on Reddit. In future work, after proposing new human and automated evaluation of topic modeling, we aim to revisit the evaluation of our proposed NTM in this chapter.

⁶⁴We discuss the potential negative impact of our work in Section 2.2 in Appendix A.14.

Chapter 3: Motivations behind Communicative Choices: Donor Activity is Associated with US Legislators’ Attention to Political Issues

In the preceding chapter, we discussed the unsupervised NLP method of topic modeling, which can be used to discover or extract themes or topics from a collection of texts. In any particular document, an author can be seen as *choosing* to talk about one or more topics out of a set of all possible topics. Considering topics again as *communicative choices*, in this chapter, we examine the particular setting of floor speeches made by US legislators (elected representatives or members of the US Congress) during a congressional session, and using associated metadata, study one possible *motivation* behind the choices the legislators make: who their campaign donors are.¹

Summary of our work and contribution. Campaign contributions are a staple of a US Congress member’s political life. Yet despite considerable research and the public’s belief that “money buys politics”, the findings are mixed. The search for tangible effects of congressional donations often focuses on whether contributions affect voting on congressional bills. We present an alternative measure by consider-

¹This chapter is based on work under review: **Pranav Goel**, Nikolay Malkin*, SoRelle W. Gaynor*, Nebojsa Jojic, Kristina Miler, Philip Resnik (2023). “Donor activity is associated with US legislators’ attention to political issues.” Working paper is available as [Goel et al. \(2023\)](#). (* indicates equal contribution).

ing the relationship between money and legislators' choices of issue-attention and prioritization as expressed through their speech. Floor speeches are an important component of congressional behavior and reflect the legislator's thinking and priorities. Our research provides the first comprehensive analysis of the association between a legislator's donors and the policy issues they choose to prioritize when speaking in Congress, discovering a significant link between the two.

We use regularized multinomial logistic regression as our machine learning framework on a new dataset that brings together legislator metadata including committee assignments, legislative speech, donation records, and information about Political Action Committees (PACs) and their industry categories. The data covers speeches made by all legislators in the House of Representatives between 1995 and 2018. We compare the predictive capacity of information about donations against other potential explanatory variables, such as party affiliation, home state, and committee assignments, and find that donors consistently have the strongest association with legislators' issue-attention. Validation by subject-matter experts confirms the discovery of meaningful donor-issue associations. We further contribute a procedure for identifying speech and donation events that occur in close proximity to one another and share meaningful connections, identifying the proverbial needles in the haystack of speech and donation activity in Congress, which may be cases of interest for investigative journalism. Taken together, our framework, data, and findings can help increase the transparency of the role of money in politics.

Our data files and replication code including expert annotations are all publicly available as a compiled repository and can be downloaded at: <https://zenodo.org/record/7465346> (DOI: 10.5281/zenodo.7465346).

Roadmap for this chapter. We will first provide the relevant background and introduce our problem statement as well as motivation in Section 3.1. We will then discuss the new dataset we create (Section 3.2) and provide an overview of our methodology (Section 3.3) as well as various details regarding the practical implementation of our models and experimentation (Section 3.4). After providing the results for our machine learning experiments that predict legislators’ issue-attention after representing them via various explanatory variables such as donors and political party (Section 3.5), we will discuss expert assessment of the donor-issue associations learned by the model during training (Section 3.6). We will then outline a procedure that can help surface potential temporal connections between speech and donation events that occur in close proximity, enabling the use of our method and its output as a tool for investigative journalism geared towards greater transparency of money in congressional politics (Section 3.7). We will conclude with the implications of our findings for political science, the limitations of our work, and how future work can build on our study (Section 3.8).

3.1 Introduction and Background

A majority of publicly-traded US-based corporations as well as many labor unions and interest groups organize a Political Action Committee (PAC) to raise funds and donate money to candidates running for political offices. PACs are major players in American politics, contributing nearly \$500 million to congressional candidates in the 2020 election cycle alone.² They play a significant role in funding congressional campaigns (Bonica et al., 2014; Hansen et al., 2015; LaPira and Thomas, 2017) and have been found to exert influence on the outcome of elections

²Source: OpenSecrets (<https://www.opensecrets.org/political-action-committees-pacs/2022>).

(Hansen et al., 2015; Kalla and Broockman, 2016; Baker, 2018). Since legislators rely heavily on donations to meet the high and rising costs of campaigns, as well as to demonstrate their electoral strength (Hassell, 2016), soliciting donations from PACs is an important and time-consuming element of legislators' day-to-day work (Torres-Spelliscy, 2017).³ There is a rich literature on PACs' decision to donate and the impact of those donations on legislators and policy outcomes. A large body of work considers the relationship between PACs and congressional elections, including the effect of incumbency advantage, differences in media costs, and voter attention (Abramowitz, 1991; Green and Krasno, 1988; Stratmann, 1991, 2009). Within the institution, prior work has considered how PACs strategically target legislators based on their existing policy positions and membership on relevant congressional committees (Hall and Wayman, 1990; Grier and Munger, 1991; Romer and Snyder Jr, 1994; Esterling, 2007; Powell and Grimmer, 2016). Other research finds evidence that donors contribute to legislators who share ideological preferences (Baumgartner and Leech, 2001; Victor and Koger, 2016; Bonica, 2018), and that donors are more likely to gain access to legislators (Hansen, 1991; Wright, 1990; Kalla and Broockman, 2016; Powell and Grimmer, 2016; Gopoian et al., 1984; Langbein, 1986; Hall and Deardorff, 2006; Baumgartner et al., 2009; Barber, 2016; Cahill and Tomashevskiy, 2019).

Research on donor activity and legislators' votes on congressional bills, however, comes to mixed conclusions regarding correlations and even less support for a causal relationship (Ansolabehere et al., 2003; Hall and Wayman, 1990; Stratmann, 2005; Roscoe and Jenkins, 2005; Smith, 1995). While there are some cases of PAC influence on specific roll-call votes (Baldwin and Magee, 2000; Goldberg et al.,

³(<https://www.cbsnews.com/news/60-minutes-are-members-of-congress-becoming-telemarketers/>).

2020), there are also numerous examples of a lack of relationship between money and votes (Bronars and Lott, 1997; Powell, 2013). The existing work likely reflects the complex reality that donations can be both an incentive and a reward, and are frequently made to legislators who tend to agree with the PAC's preferences, all of which makes isolating the causal impact of money difficult. Even without definitive findings, this body of research raises concerns about the potential role of money in politics, particularly given the bias towards economic organized interests (Schattschneider, 1960; Schlozman et al., 2012).

Identifying reliable associations is the starting point to understanding potentially important relationships between donations and congressional behavior. Towards this goal, scholars have turned to examining the impact of money on behaviors beyond votes. While roll call votes constitute an important legislative action, they are limited in their ability to capture the full range of positions and priorities of individual legislators, and thus perhaps less affected by donors. Rank-and-file members have little autonomy over the congressional agenda, and the binary (yes-or-no) nature of roll call votes limits the positions that votes can convey. In an increasingly centralized and partisan legislative environment, members are often forced to “toe the party line” and lose the chance to carve a separate position for themselves either within their party or in the congressional chamber as a whole (VanDoren, 1990; Harris, 2005; Carrubba et al., 2006). Conversely, legislators have much more discretion over their participation in other parts of the legislative process as well as their *rhetoric* (Diermeier et al., 2012; Grimmer, 2013; Grimmer and Stewart, 2013). For instance, scholars have found evidence that donors affect legislators' behavior in committees (Hall and Wayman, 1990; Hojnacki and Kimball, 1998; Esterling, 2007; Powell, 2013). More recent work has begun to investigate the relationship of donors to legislative rhetoric, including how donors influence the

adoption of suggested legislative text (Hertel-Fernandez, 2019; McKay, 2018), the strategic use of interest groups' committee speech to influence legislators' policy positions (Vallejo Vera, 2023), and the bidirectional relationship between donors and US senators when examining committee discussions on energy policy (Iliev, 2021).

Our work contributes a new avenue to the study of the association between money and non-voting behavior by focusing on the issue agenda in Congress and the specific issues that legislators choose to call attention to on the House floor. Floor speeches allow individual legislators to convey proximity to party leadership or distance themselves from the party "message" on any policy (Harris, 2005). And while there are some limitations when it comes to scheduling and time allotted to speak,⁴ floor speeches provide a rich view of a legislator's thinking and priorities (Lauderdale and Herzog, 2016; Vafa et al., 2020) via the attention allocated to various policy issues (*issue-attention*) (Witko et al., 2021) and framing (Nguyen et al., 2015b).

Legislators' issue-attention is tied to the agenda-setting function of communication and rhetoric by the political elite. Prior work has considered issue-attention *i)* a direct proxy of issue prioritization by policymakers (Jones and Baumgartner, 2005), *ii)* an important criterion for driving policy action both inside Congress (Baumgartner and Jones, 1993; Maltzman and Sigelman, 1996) as well as outside Congress (Eshbaugh-Soha, 2005), *iii)* an influence on public opinion (Broockman and Butler, 2017), and *iv)* a powerful aspect of wielding political power, especially when attention influences which issues become or do not become a part of the policy agenda in Congress (Bachrach and Baratz, 1962). This can be of great importance to underrepresented groups in Congress (Witko et al., 2021; Osborn and Mendez, 2010;

⁴Members cannot explicitly campaign on the House floor or use floor speeches directly in campaign materials. The House Committee on Ethics determines and enforces these limitations.

Pearson and Dancey, 2011a). Floor speeches are an avenue for members of Congress to proactively show their expertise on policy issues (Burden, 2007), increase their visibility, and demonstrate their commitment to particular issues and stances — not just to their fellow representatives but to the press and their constituents (Highton and Rocca, 2005; Rocca, 2007; Cormack, 2016) as well as outside political actors including PACs. With reelection-minded legislators looking to communicate their preferences to their constituents (Mayhew, 2004), floor speeches are an important tool for legislators to convey how they represent their constituents and to create a particular public image (Grimmer, 2013) that is often not possible through roll-call votes. Legislators can also use speeches to signal to PACs and interest groups their shared policy preferences and willingness to advocate on those issues. Our work can therefore also help document the professional relationships that PACs and legislators seek to build (Hall and Deardorff, 2006; Kroszner and Stratmann, 2005).

The general importance of PAC donations to a legislator’s political career, as well as the importance of floor speeches as a critical legislative action that offers insights into congressional agenda setting, motivates us to conduct the first examination of the association between PAC donations and the attention placed by legislators (as the recipients) on various issues in their floor speeches.⁵ Although this research remains agnostic regarding causal directions, we uncover a strong and robust association between money and legislative speech, which reveals that money has a more extensive reach in American politics than previously established and proposes a foundation for future work to investigate the potential causal relationship.

⁵To the best of our knowledge, this potential link between donations received by legislators and their floor speeches has not been studied in prior literature, although such an investigation has support in related studies on the influence of money on language in other contexts. Our research greatly expands the scope of analysis by looking across congressional sessions and policy issues, and considers all legislators in a congressional body instead of a subset as well as all of the floor speeches as opposed to a particular proceeding.

In this work, we seek to answer the following question: **are donors substantially associated with legislators’ issue-attention when speaking on the floor of Congress?** To study this, we compare the association provided by information about donors with other information about legislators such as their home state, party affiliation, and committee assignments (that serve as alternative explanations of legislative decision-making).⁶ We find that donor information consistently offers the strongest association with legislators’ issue-attention. This uncovers a previously unidentified relationship and provides new evidence of how money and issue priorities go hand-in-hand. This association can help inform our understanding of legislative motivations and enable greater transparency and accountability of elected officials.

In addition to our main finding that donor information is meaningfully associated with legislators’ issue-attention in their floor speeches (both in the aggregate and across individual congressional sessions), we gain new insights into PAC dynamics by uncovering both the top donors within each issue area and the top issues for each industry.

We create a new dataset that includes information and metadata about legislators, along with the text and dates of floor speeches, donations they receive from PACs, and information about the PACs such as the industry they belong to.⁷ Our comprehensive dataset covers the US House of Representatives over twelve congressional cycles, from 1995 to 2018 (that encompasses periods of both Republican and Democratic control).⁸

⁶In particular, we note that while all representatives request particular committee assignments in the US House, some get their preferences while others do not; there is an influence of committee assignment on legislators’ behavior in either case via different mechanisms (Miler, 2017).

⁷We note that the committee assignment data we use includes select as well as joint committees and the various modified committees created over the years covered in our dataset (that is made publicly available).

⁸The US Supreme Court’s ruling in *Citizens United v. FEC* (2010) affected campaign finance

Using our new dataset, model, and the expert validation of issue-donor associations given by the model, we also introduce a procedure that highlights specific donation and speech events that *a)* occur in close proximity to each other, and *b)* stand out as significantly connected (*i.e.*, are more likely to *not* constitute a random occurrence of a pair of events close in time). This procedure can find cases where a donor donates a significantly large amount to a legislator within days (before or after) of the legislator giving a floor speech on a policy issue that the donor cares about (an example is provided in Fig. 3.8). We identify these cases considering every congressional cycle in isolation and can score and rank them based on the strength of the association.⁹ Experts and journalists can then use the filtered set of connected donation and speech events occurring within a small time window (less than a month; before or after) as starting points to identify situations that potentially warrant further investigation. Our codebase documents this procedure and the ranked cases. Our dataset, model, and analyses will enable further work on the connections between legislators’ language use and their donors.

3.2 Dataset Overview

Raw data sources. We briefly discuss the sources of our various raw data components, our process for creating a new dataset, and provide an overview of its contents.

We obtain raw transcripts of **floor speeches** given by legislators in Congress (along with some associated legislator information) from the US Congressional

law in the area of outside independent spending in federal campaigns. The ruling gave rise to independent expenditure-only PACs (also called Super PACs), but these are distinct from the type of traditional hard-money PACs examined here. Traditional PACs donate money directly to candidates’ campaigns and were not the focus of the Citizens United case.

⁹Our procedure is discussed in Section 3.7.

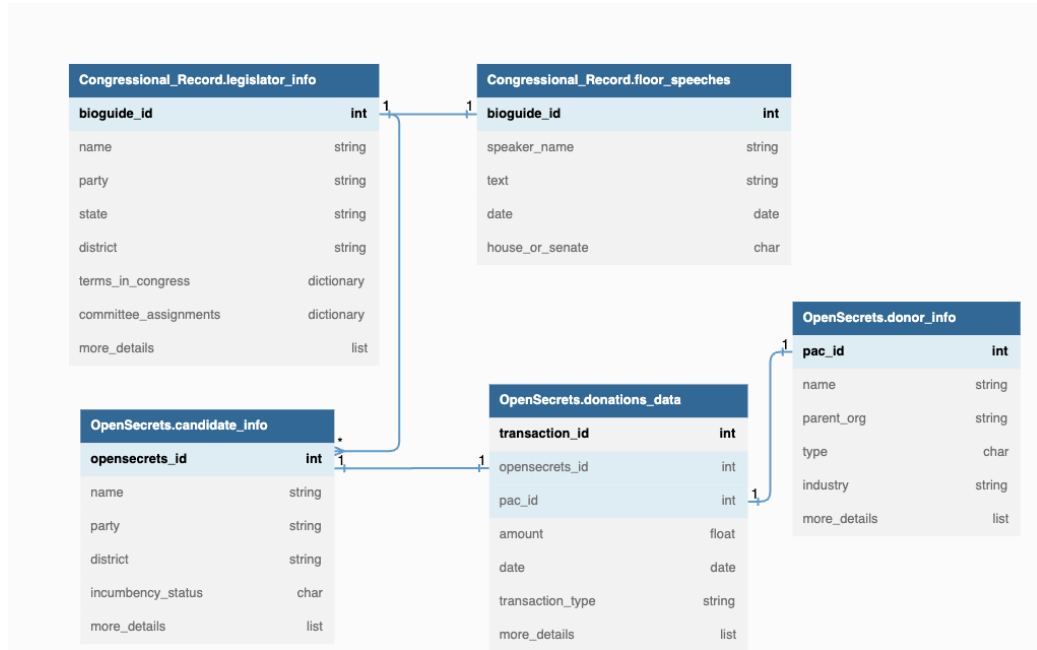


Figure 3.1: Our new dataset’s schema. Our main data contribution relies on creating a mapping between the *bioguide_id*, which is the primary key or ID for legislators in congressional Record Data providing floor speeches transcripts as well as the repository providing the accompanying biographical and committee membership information), and the *opensecrets_id*, which is the primary key or ID used to refer to political candidates receiving donations in the OpenSecrets database providing campaign contributions transactions data. Note that the columns or information shown for each table is not exhaustive; this is just an overview of the structure and contents of our database. The *pac_id* is originally found in FEC data and used in the bulk data made available by OpenSecrets that we use. There are 5 possible values for the type of PAC: Business, Labor, Ideological, Other, and Unknown — we only consider Business and Labor PACs in our study. All the various ID mappings in our database are either one-to-one or one-to-many, represented using ‘1’ and ‘*’ on the mapping links in the schematic diagram.

Record database (Judd et al., 2017).¹⁰ We were able to obtain floor speeches from 1995 to 2020.

For comprehensive **information about legislators** such as their state and district, party, the terms they serve in Congress, biographical details, and their committee and subcommittee assignments, we use the public repository at <https://github.com/unitedstates/congress-legislators>.

¹⁰<https://github.com/unitedstates/congressional-record>

Raw data on US political campaign **donations** is made available by public governmental databases like the Federal Elections Commission (FEC).¹¹ OpenSecrets is a nonprofit research organization that tracks donations and makes available bulk data on donation transactions as well as donors; we use their publicly available donations database in this study.¹² Using the bulk dataset, we obtained campaign contribution transaction data for US legislators from 1989-2018.

In order to work with a connected donations and floor speeches dataset, we utilize information over the 1995-2018 period, covering 12 congressional sessions or cycles.

Process for linking different available databases in order to create our new dataset.

The schematic diagram and brief overview along with some details of our new database are provided in Fig. 3.1.

A common primary ID used in all the relations included in the US Congressional Record database (containing floor speeches and legislator information such as their state, party, committee assignments, etc.) is the 'BioGuide ID' (`bioguide_id` in Fig. 3.1). In the donations database made available by OpenSecrets, the recipients of these donations are primarily identified via an ID, which we call the 'Opensecrets ID' (`opensecrets_id` in Fig. 3.1).

To link together these databases in order to enable research that can link legislator information and rhetoric with donations, we map the OpenSecrets ID and BioGuide ID for legislators who were part of the US House over the 1995-2018 time period, first using available data on legislators available in both databases to automatically connect the IDs using exact matching (the whole name as well as other shared information such as party and date of birth) to ensure complete precision, and then

¹¹<https://www.fec.gov/>

¹²<https://www.opensecrets.org/bulk-data/>

perform manual mappings in cases where the automated approach did not find an *exact* match. The mapping is verifiably exact.

This reliably maps the two IDs in order to enable the analyses in this chapter, and by making all our data publicly available, we hope that our efforts will help future research as well.¹³ In this work, we focus on the US House, in terms of the legislators as well as the floor speeches.

Note on industry-based groups of PACs. We also use another resource on PACs (donors) — manually coded group assignments based on the sector, industry, and the specific industrial category that the PAC belongs to as obtained by OpenSecrets.¹⁴ This is a hierarchical coding scheme that provides for each PAC the very broad sector (such as transportation) it belongs to, its industry within that sector (such as air transport), and its specific category within that industry (such as air transport manufacturing). Such industry information is a handy reference for contextualizing the relevant policy area for PACs (and, therefore, useful for validation of our model outputs; Section 3.6). We also test these categorical groups of PACs for their association with legislators’ issue attention (Section 3.4).

Note on types of PACs in our processed dataset. Our processed dataset consists of both business (or corporate) and labor PACs. Post-processing, we have 1002 PACs in total (Table 3.1). Out of these, 937 are business PACs and 65 are labor PACs. Business PACs donate to more legislators on average in our data, and also have a significantly higher association with issue-attention than labor PACs (detailed results and analysis are provided in Appendix B.2). This novel examination adds to

¹³Our novel dataset shares some similarities with some existing database creation efforts (Bonica, 2019, cf.).

¹⁴<https://www.opensecrets.org/PACs/list.php>

prior work that also focuses on the critical dynamics of the influence of business and labor money in American politics ([Bauer et al., 1963](#); [Sorauf, 1988](#); [Box-Steffensmeier et al., 2005](#); [Witko et al., 2021](#); [Hertel-Fernandez, 2014, 2018, 2019](#); [Drutman, 2015](#); [Frymer and Grumbach, 2021](#)).

We provide the dataset processing details as well as statistics of the processed dataset used in our study (including the number of speeches, legislators, donors, and sizes of other categorical variables) in Section [3.4.1](#).

3.3 Methodological Framework

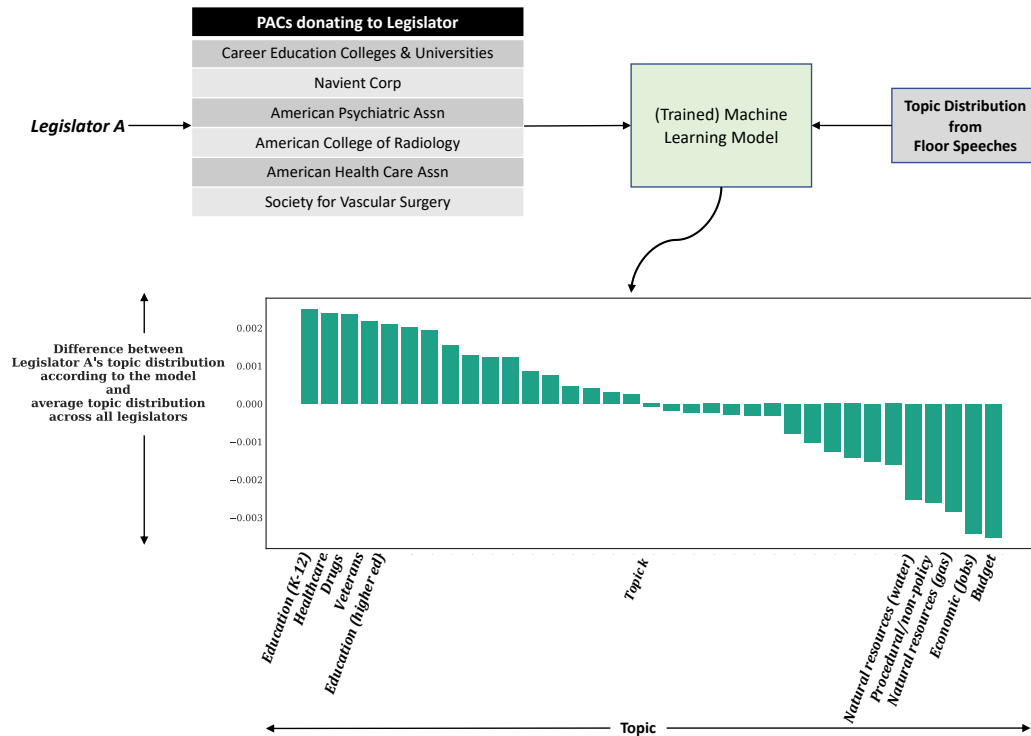
We provide an overview of our method using an illustrative example with real data (and our actual trained model itself) in Fig. [3.2b](#).

We start by considering information about legislators and the content of all their floor speeches over the entire period of our data (1995-2018). Issue-attention is quantified by running a topic model ([Blei et al., 2003](#)) on all the floor speeches to identify a collection of K topics or themes; each speech can then be considered a mix of the discovered topics.¹⁵ The corresponding value for a given topic in the probability distribution over topics for a speech indicates the presence of that topic in that particular speech. For each legislator, we then take the mean of these topic probability values over all their speeches to capture how much they talk about each topic on average.

¹⁵Using discovered issues (topics) from text as a key component in our analysis has some support in previous work that studied PAC influence on roll call votes and committee deliberation in the US Congress and found that the manner of the PAC influence is linked to the type of issue and the issue context ([Witko, 2006](#)). Furthermore, corporate funding has been found to influence the themes present in the discourse around climate change ([Farrell, 2016](#)), and thus, studying the association between donations and issue-attention in text is a viable framework.

Topic (top terms)	Label (Expert 1)	Label (Expert 2)
education, schools, students, teachers, districts, parents, kids, teacher, quality, educational	Education (K-12)	Education
medical, patients, doctors, hospital, patient, hospitals, doctor, quality, providers, physicians	Healthcare	Health care
drug, drugs, tobacco, abuse, colombia, fda, substance, illegal, marijuana, addiction	Drugs	Drugs
veterans, va, benefits, affairs, deserve, receive, disabled, disability, disabilities, families	Veterans	Veterans
education, college, students, programs, training, student, workforce, institutions, opportunities, loans	Education (higher ed)	Education

(a) Examples of issue labels created independently by two experts. Note that experts were provided both the top terms and top documents per topic modeling output to create labels for each topic.



(b) An illustrative example of our approach and what it captures, using a simulated legislator and real data. For PACs shown as donors for the individual simulated 'Legislator A', we show the difference between the average topic distribution across legislators (derived from all the floor speeches) and Legislator A's topic distribution per our trained machine learning model that represents legislators using which PACs donated to them. "Navient Corporation is an American student loan servicer based in Wilmington, Delaware" (source: <https://en.wikipedia.org/wiki/Navient>). Issue labels provided by one of our experts (expert 1) are used for topics shown on the x axis.

Figure 3.2: Our model automatically quantifies disproportionate attention paid by legislators to the issues relevant to the PACs donating to them — this is illustrated here using a simulated legislator as an example, using the actual average topic distribution across all legislators (and their speeches in our data) and the legislator's topic distribution per our model. With healthcare and education-related PACs donating to the legislator, the model trained to capture associations between donors and issue-attention expects a disproportionate focus on education and healthcare topics.

To quantify the association between legislators' issue-attention and information about the organized donors who donate to them, and to evaluate potential relationships between speech content and donation, we use a supervised machine learning approach using logistic regression. In brief, descriptive information is coded as a vector of binary variables for each legislator: 1 if that particular category applies to them and 0 if not (one-hot encoding). We refer to all these various ways to represent information about legislators as *legislator attributes* (or representations), and these are the independent variables in the logistic regression setup.

The legislator attributes can then be used to predict the topic values or issue-attention, which is the dependent variable in the logistic regression model, and the comparison between the predicted issue-attention and the actual values helps determine the extent to which the legislator attribute-based prediction explains the known issue-attention derived from their floor speeches. This quantification of the explanation or level of association is then compared for different legislator attributes, which include: *PAC* (binary vector based on if each donor (PAC) donated to the legislator or not), *Committee* (based on the legislator serving on particular House committee or not), *State*, and *Party* (all are one-hot vectors encoding presence or absence of various possible values). Each attribute can have a number of different values (such as Republican, Democrat, or Independent for the *Party* attribute)), and we refer to the total number of possible values for any attribute as the size of the attribute (or the dimensionality of the particular representation).

To understand the impact of the coarseness of the information on donors, we compare representing a legislator based on donations by *individual* PACs with a representation based on donations made by a *group* of PACs. These constitute two additional, more coarse-grained, donor-based attributes for legislators: one examining the effect of legislators receiving donations from a specific industry (*e.g.*,

‘Public Unions’), which we refer to as the *Industry* attribute, and the other examining the effect of legislators receiving donations from a particular subset of the industry (e.g., ‘Teacher Unions’), which we refer to as the *Category* attribute.

We have 1002 individual PACs in our final processed dataset, while other representations (or attributes) have fewer dimensions (or sizes/possible values), such as 3 for *Party* (Table 3.1). Because of the order of magnitude difference in sizes of the different legislator attributes, we include an additional legislator attribute as a baseline for comparison called *Random-PAC*, where the *PAC* attribute is ‘shuffled’ so that each legislator has a binary encoding over PACs for a random, different legislator (each legislator is being represented by donor information of some other legislator). This helps assess if any difference in the predictive power of the *PAC* attribute and other attributes is simply due to the size of the *PAC* attribute and also serves as a null hypothesis for information on PACs or donors having a significant predictive effect for issue-attention. Put another way, if *Random-PAC* could predict issue-attention as well as *PAC*, that would indicate that knowing which particular PACs donated to a legislator under consideration does not have any association with their issue-attention.

To assess the predictive strength of a particular legislator attribute, we use regularized multinomial logistic regression model and use cross-validation; measurement of the strength of association is enabled by computing the reduction in bits required to encode issue-attention (based on the information-theoretic concept of optimal encoding of probability distributions). Details about data processing, the model, training, and measurement are provided in the next section (Section 3.4).

In addition to training and using one topic model and logistic regression model for the entire data spanning twelve congressional cycles, we also train separate topic models and regression models for each congressional cycle. The process is

Speeches	Legislators	PAC	Industry	Category	Committee	State	Party
386532	758	1002	72	250	94	56	3

Table 3.1: The number of speeches and legislators in our processed dataset, along with the sizes of each of the legislator attributes used as possible explanatory variables in our study (*Random-PAC* is the same as *PAC*). The attributes (or legislator representations) are described in Section 3.3. These details are applicable to our model trained on the entire dataset; we provide these attribute sizes for models trained separately on data from each congressional cycle in Table B.2. We note that the six other possible values for *State* are: American Samoa, District of Columbia, Guam, Northern Mariana Islands, Puerto Rico, and Virgin Islands.

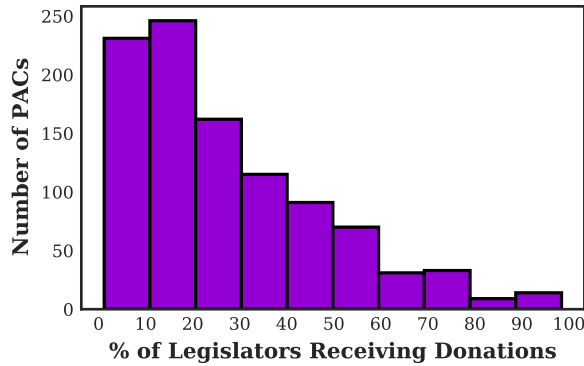


Figure 3.3: In our processed dataset covering 758 selected legislators giving floor speeches across 1995-2018, and 1002 PACs actively donating in the same period; most PACs tend to donate to less than a third of all the legislators they could donate to.

largely the same with a few training details and choices changed (Section 3.4.3). This enables the comparison of the same legislator attributes highlighted above in terms of their predictive capacity for issue-attention of US House representatives over the time period captured in our dataset.

3.4 Experimental Detail

3.4.1 Data Processing

We use the speeches, metadata, and donation information for the members of the US House over the period from 1995 to 2018, covering twelve congressional

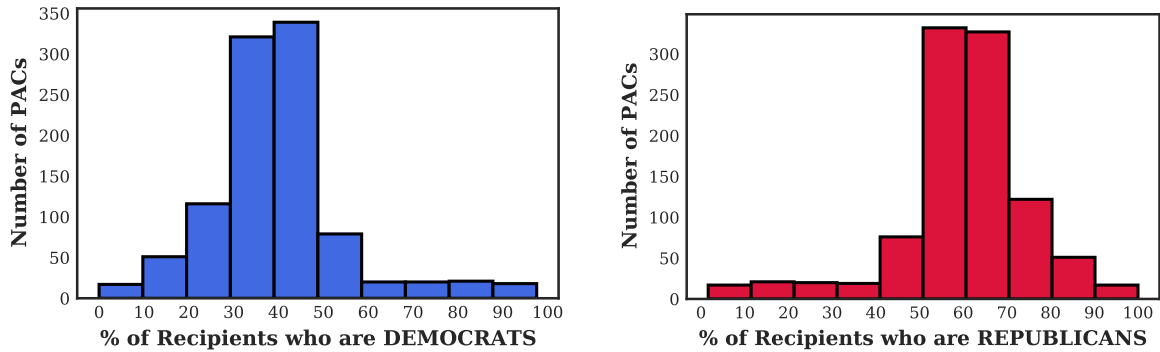


Figure 3.4: In our processed dataset used to train our machine learning framework, most PACs tend to donate across party lines instead of donating in an exclusive, partisan manner.

cycles. We retain representatives who gave at least fifty floor speeches and gave at least one speech in six out of the twenty-four years in the time period covered in our data. We retain donors (PACs) who donated at least \$1000 to at least five of the selected representatives in at least four out of the twelve congressional cycles in our data. This gives us the final data statistics presented in Table 3.1 that are used in our logistic regression model (trained over the entire dataset).

We use similar filtering when the models are trained separately for each congressional cycle: legislators giving at least five speeches and being recipients of at least three different donation transactions (in a particular congressional cycle) are retained for that cycle. Donors making at least three different contributions in that cycle are retained for that cycle. The final statistics for the dataset used in model training for each individual cycle are presented in Table B.2.

In our processed dataset, most PACs donate to fewer than a third of legislators and therefore are somewhat targeted in their donations (Fig. 3.3). In addition, most PACs in our data are not donating in an ideological or clearly partisan fashion; instead, they donate to a mix of Democratic and Republican legislators (Fig. 3.4).

3.4.2 Topic Modeling

To quantify how much each of the different issues is talked about in the set of speeches given by a legislator, we use topic modeling, specifically, LDA (Blei et al., 2003) estimated with Gibbs Sampling (Griffiths, 2002) as implemented in the MALLET package (McCallum, 2002).¹⁶

As explained in Chapter 2, LDA models a collection of documents through a collection of latent topics or themes. Each topic is represented as a probability distribution over the words in the vocabulary, and each document is represented as a mixture of topics or themes. This is an unsupervised method, and by looking at the top words and the top documents assigned to a topic, we can interpret what the discovered theme is about or the issue it represents. Two political scientists independently labeled each topic in our work.

The instructions used for obtaining labels for topics and annotations are provided in full as part of our codebase and data. A few examples of topics and labels are included in Fig. 3.2a (more are provided in Appendix B.4).

When looking at the predictive capacity of various legislator attributes for issue-attention over the entire set of floor speeches from 1995-2018 considered together, we run LDA with number of topics $K = 60$. We also run a separate 30-topic model on floor speeches made within each of the twelve congressional cycles (1995-96 to 2017-18) in our data (lower K due to fewer speeches in an individual cycle compared with the entire dataset), which is used to train separate models (Section 3.4.3) and helps us focus on a two-year time period at a time in order to analyze how the results change over time across congressional sessions.

¹⁶Hoyle et al. (2022) found classical LDA to be the dominant topic model of choice among practitioners and to be better than newer neural topic modeling methods on criteria grounded in the qualitative content analysis use case; and in Section 2.2, we found that LDA yielded strong qualitative ratings for its topics as judged by humans.

We note that while experts tend to come up with the same policy or issue labels for topics, sometimes there are differences in the level of specificity between the two experts — for example, ‘education’ versus ‘higher education’.

Robustness checks for testing dependency on particular topic modeling output.

As a critical robustness check, especially considering the unsupervised nature of topic modeling, we provide results after non-meaningful, non-policy-related topics per the judgment of either of the two experts are removed in Appendix B.2. We observe no change in our findings and interpretation of the results. We explicate the labels that are considered non-meaningful or non-policy-related in Appendix B.4.

Furthermore, we also conduct our main experiment with legislator issue-attention quantified using an entirely different topic model: the Structured Topic Model (STM, Roberts et al., 2013, 2014). Our main finding and pattern based on the document-topic posterior obtained by STM precisely mirrors our findings using LDA (Appendix B.2). Therefore, the findings in this work are not dependent on automatically discovered topics alone, but in fact, are reliably established across both the initial set of automatically discovered topics as well as the expert-refined or curated set of issues.

3.4.3 Regularized Multinomial Logistic Regression Framework and Training Detail

We discussed the process for obtaining the average topic distribution for every legislator as well as the various explanatory variables or legislator attributes above in Section 3.3.

Let the matrix of averaged topic probabilities (number of legislators by number

of topics) be y (dependent variable) and a given legislator attribute be X (independent variable). To predict y from X , we train a L_2 regularized multinomial logistic regression model:

$$y' = P(X) = \frac{1}{1 + \exp^{-(\beta_0 + \beta_1 X)}} \quad (3.1)$$

β_0 is the bias or intercept term, and β_1 is the weight term (together, these are the regression coefficients of our model). y' is the predicted topic probability distribution for the K topics, and it can be compared with the actual topic distribution derived from floor speeches in order to train the model. If the dimensionality (or size) of the independent variable X is m , then the weight matrix β_1 will be of the size $K \times m$ (for e.g., 60×1002 in case of the legislator attribute being *PAC*). This weight term of the trained logistic regression model indicates the learned associations between each of the values of an attribute (such as an individual *PAC*) and each of the topics.

To compare legislator attributes in terms of their association with issue-attention in floor speeches, we train our logistic regression model on some legislators to predict their averaged topic proportions, and then check the learned predictive capacity on a held-out set of legislators. Specifically, we use 50-fold cross-validation, which means 2% of legislators are randomly selected to be in the held-out set for predicting their topic distribution while the model trains on the other 98% legislators to closely approximate their given topic distribution; and this is repeated 50 times. This kind of predictive framework tests the various ways to represent legislators for their capacity to generalize, i.e., given a new set of legislators and their donor, party, state, or committee information, we test how well the topic distribution predicted from that information matches the actual topic distribution for those legislators.

The multinomial logistic regression model is trained using the cross-entropy loss using the target distribution, which is the topic distribution for each legislator.

When training a model for one particular congressional cycle at a time, the process is the same, except a 30-fold cross-validation procedure is used to account for the fewer legislators present in a given single cycle.

The process for fine-tuning and selecting the hyperparameters of the logistic regression model, such as the learning rate and the L_2 regularization coefficient, along with the final optimal settings used for training and cross-validation of regression models for each of the legislator attributes are provided in Appendix [B.1](#).

3.4.4 Measuring the Association

In our framework, the regularized multinomial logistic regression model is trained to predict a probability distribution vector with K values (where K is the number of topics) by learning an association between a legislator attribute or representation and given probability distributions on the training set. This training relies on optimizing the value of cross-entropy between the predicted and actual topic distributions on the held-out or validation set (this process is repeated as per the cross-validation procedure).

To measure the association between a legislator attribute and issue-attention, we recognize that the machine learning setup converts legislator attributes into ‘predicted’ issue-attention distribution, which can be compared with the actual issue-attention distribution derived from floor speeches. Since we are dealing with probability distributions, we use the information-theoretic concept of optimal encoding of distributions for our measurement.

Consider two sets of legislators: ones in the training set, and ones in the held-out

set. The goal for our machine learning method trained using a particular legislator attribute (as the ‘model’) is to predict or provide its best estimate of issue-attention on the held-out set. The expected lowest number of bits required to encode issue-attention for the held-out set is the negative entropy of the distribution. This provides us with the *lower-bound*. Note that a lower number of bits is better since it implies a more efficient (or optimal) encoding of a distribution.

In the absence of any information on legislators, we can use the issue-attention available in the training set, and measure the cross-entropy between the training set and the held-out set issue-attention (topic distributions) in bits of information. This provides us with an *upper-bound*. In the absence of information on legislators (any of our legislator attributes), the maximum distance (in terms of bits of information) we could reduce is given by the difference between this upper bound and the aforementioned lower-bound.

In the *presence* of information about legislators, such as which PACs donated to them, we can use our machine learning method to obtain an estimate for the held-out issue-attention. Cross-entropy once again measures the bits required to encode the actual held-out issue-attention given the estimated issue-attention provided by a model of the legislator. We call this the *model cross-entropy*. Our measurement (used to obtain our results in Section 3.5 below) is called the % **bit reduction**, and is computed as follows:

$$\% \text{ Bit Reduction} = 100 * \frac{\text{upper-bound} - \text{model cross-entropy}}{\text{upper-bound} - \text{lower-bound}} \quad (3.2)$$

Ultimately, our measurement computes the reduction in bits required for encoding issue-attention — a higher reduction implies that a more optimal encoding is enabled by relevant information about legislators. This can also be framed as the

particular legislator attribute providing more explanation for the variation in the issue-attention of the legislators. This measurement helps us both compare various legislator attributes and quantify the explanation offered by the attribute.

3.5 Results

PACs are significantly more associated with legislators' issue-attention in floor speeches than other legislator attributes. Considering information about legislators and the content of all their floor speeches over the entire period of our data (1995-2018), Fig. 3.5a shows that knowing which PACs donated to a legislator explains more of the variation in legislators' issue-attention than knowledge of other legislator attributes. This effect is statistically significant ($N = 50$, $p < 0.05$; details provided in Appendix B.2) and reveals that *PAC* information provides the strongest association with legislators' issue-attention. Other non-donation-related attributes about legislators, such as their party, home state, and committee assignments are significantly less predictive of the issues that legislators talk about than PAC donations. This is a striking finding because one expects parties, committees, and state interests to be tightly related to the issues that legislators spend their time talking about in the House. Notably, information about individual PACs (such as 'Career Education College & Universities' PAC) is more predictive than information about industry-based groups of PACs (such as the *Industry*: 'Education'; or even the specific *Category* within the industry: 'For-profit education').

Finally, we find that the effect is not simply due to the higher number of possible values (or size) for the *PAC* attribute as shown by the baseline attribute of *Random-PAC*, where the donor profile of legislators is randomly shuffled. We note that the negative result for *Random-PAC* is due to the machine learning model *overfitting*

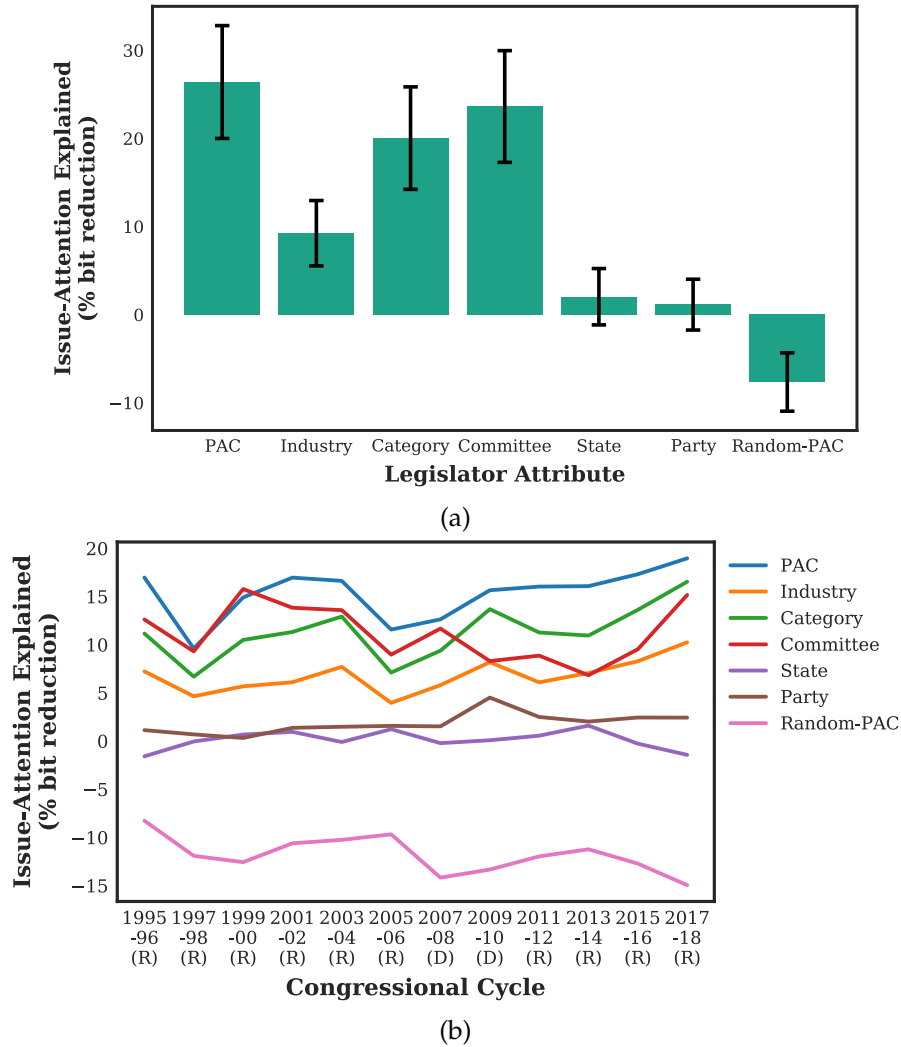


Figure 3.5: PACs are most associated with legislators' issue-attention in floor speeches compared with other legislator attributes, both in aggregate terms across the entire time period of our dataset as well as in individual congressional cycles. (a): shows the results in aggregate, with information about which individual PACs donated to legislators clearly explaining their issue-attention the most out of all the legislator attributes we consider; the cross-validation validation procedure was repeated 50 times, and the error bars (for results across the cross-validation procedure) are shown. (b): shows the results when repeating our topic modeling and machine learning procedure separately for each of the twelve congressional cycles in our data, and demonstrates the same pattern as the aggregate picture (with the cross-validation procedure repeated 30 times; the error bars are provided in Fig. B.3). (R) and (D) indicate which party was in the majority in the US House during that congressional cycle. Both of our findings are statistically significant (Appendix B.2).

to the training data. In other words, the information did not allow the model to generalize in its predictions of issue-attention for legislators not seen during the training of the model (the held-out set, as discussed above in the method overview). This ‘overfitting’ effect for *Random-PAC* is made clear by viewing the corresponding results on the training set provided in Appendix B.2.

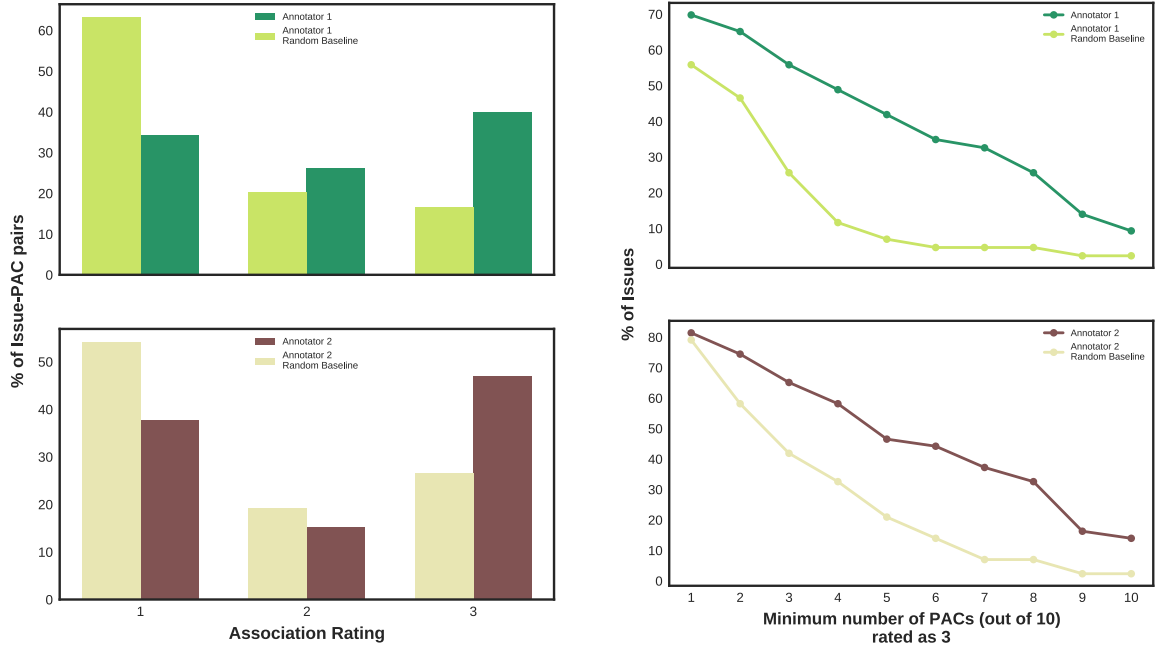
PACs consistently offer the strongest association with issue-attention across all congressional cycles. As we mentioned above in Section 3.3, in addition to training and using one topic model and logistic regression model for the entire data spanning twelve congressional cycles, we also train separate topic models and regression models for each congressional cycle. The same legislator attributes can then be compared in terms of their predictive power for issue-attention of US House representatives.

We find that *PACs* are the best predictive indicator for every cycle (Fig. 3.5b).¹⁷

We note that conducting our experiment separately for each congressional session also allows us to quantify another legislator attribute: legislator *seniority*, based on the number of congressional terms the legislator has served up to and including the particular congressional session being analyzed. Results for this specific attribute are noted separately in Appendix B.2 and conform to the aforementioned finding; indeed, seniority is unable to explain issue-attention with any significance.

In both cases — models trained on the entire data, and separate models trained for each individual cycle in the data — apart from donation information (*PAC*), the only other competitive legislator attribute appears to be knowledge about *Committee*

¹⁷The difference between predictive strength of *PAC* compared to other attributes is significant for most cycles ($N = 30$, $p < 0.05$, details in Appendix B.2).



(a) Annotators rated PACs shown for an issue on a 3-point Likert scale (3: the PAC and the issue are related, 1: they are unrelated). The top PACs for each issue per our model are rated higher than a random selection of PACs for both our annotators (rating independently). The difference between the number of issue-PAC associated rated as related (3) when PACs shown are as per our model versus a random selection is statistically significant (Appendix B.2).

(b) A value of 5 (as an example) on the x-axis means *at least* 5 out of the 10 PACs shown were rated as clearly associated with the issue (a rating of 3). More PACs were rated as related to the issue, out of the 10 shown, when the PACs shown were selected based on the associations learned by our model compared with a random selection.

Figure 3.6: Results of annotating PAC-issue associations, showing that the top PACs per issue as per our regression model’s weights are meaningful since the model’s top 10 PACs are rated consistently higher in terms of association with the issue as compared with a random selection of 10 PACs for the issues.

assignments (Figs. 3.5a and 3.5b). We, therefore, explore the comparison for *PAC* and *Committee* in greater detail in Appendix B.3 and also show that combining the two attributes offers a higher bit reduction than using just one or the other, which indicates there is at least some complementary predictive capacity for issue-attention present in these two sources of information on the legislators.

3.6 Expert Assessment of PAC-Issue Associations Learned by Our Model

The topic model identifies themes or categories present in a collection of documents in an unsupervised manner but does not create a meaningful label for the discovered theme by itself. A probabilistic topic model (such as LDA) assigns values for each topic for every document as well as for each term in the vocabulary for every topic. Two political scientists independently reviewed the top terms and the top floor speeches associated with a topic and assigned labels based on their understanding of US politics. This allows us to move from *topics* (automatically discovered themes or categories) to meaningful political *issues* or issue-labels (curated by humans working with the output of a topic model) in our analyses. Having these labeled issues helps further analyze and understand what our model learns when working with data on the PACs donating to legislators.

The weight parameter learned by our logistic regression model (β_1 in Eq. (3.1)) indicates the learned association between each dimension of information used to represent the legislator (*e.g.*, each particular PAC, committee, state, party, etc.) and each dimension in the dependent variable being predicted (*i.e.*, each topic discovered by the topic model trained on all the floor speeches). By sorting the weight values of the trained model, we obtain the top 10 PACs (representing $\sim 1\%$ of all PACs in the processed data) for every issue.¹⁸ Some examples of issue labels for topics created by the two annotators or subject-matter experts are shown in Fig. 3.2a. More examples of issue labels as well as the top 10 PACs for some issues per the model are provided in Appendix B.4.

¹⁸We only consider meaningful topics here as issues per the labels created by political scientists; the discarded labels are provided in Appendix B.4.

We then create an annotation ask, independently asking the two political scientists to rate each of the PACs shown for the issue (presented with the label they created) as *Related* (3), *Potentially related* (2), or *Unrelated* (1). To judge whether the ratings for top PACs for issues as provided by our model are more meaningful than we would otherwise expect, the same task was repeated with the same setup except the 10 PACs shown for every issue were selected randomly from all the PACs in our data (the annotators were not privy to this difference between the two setups – they received the two tasks with the same instructions). We refer to this latter setup as the ‘Random Baseline’.

Apart from the names of the PAC, information about the sector, industry, and specific industrial category they belong to are also contained in the same annotation spreadsheet to provide additional context about the PAC. Annotators were free to search for additional information about the PACs if they needed to before providing a rating, but were never required to do that.

More details about the annotation task for validating PAC-issue associations including instructions and examples of what the annotation task looks like are provided in Appendix B.4, and the final annotations are provided along with our raw data, processed data, and modeling outputs in our publicly available codebase.

Figs. 3.6a and 3.6b show that the PACs found to be most associated with issues by our model are deemed as meaningfully associated at a higher rate than for randomly selected PACs for those same issues; we find that this difference is statistically significant (Appendix B.4). This shows that our approach enables a large-scale probabilistic analysis that does uncover real, validated signals in the big donations and unstructured language data.

In summary, we show that the PAC-issue associations learned automatically by the model during its training in order to predict issue-attention are meaningful in

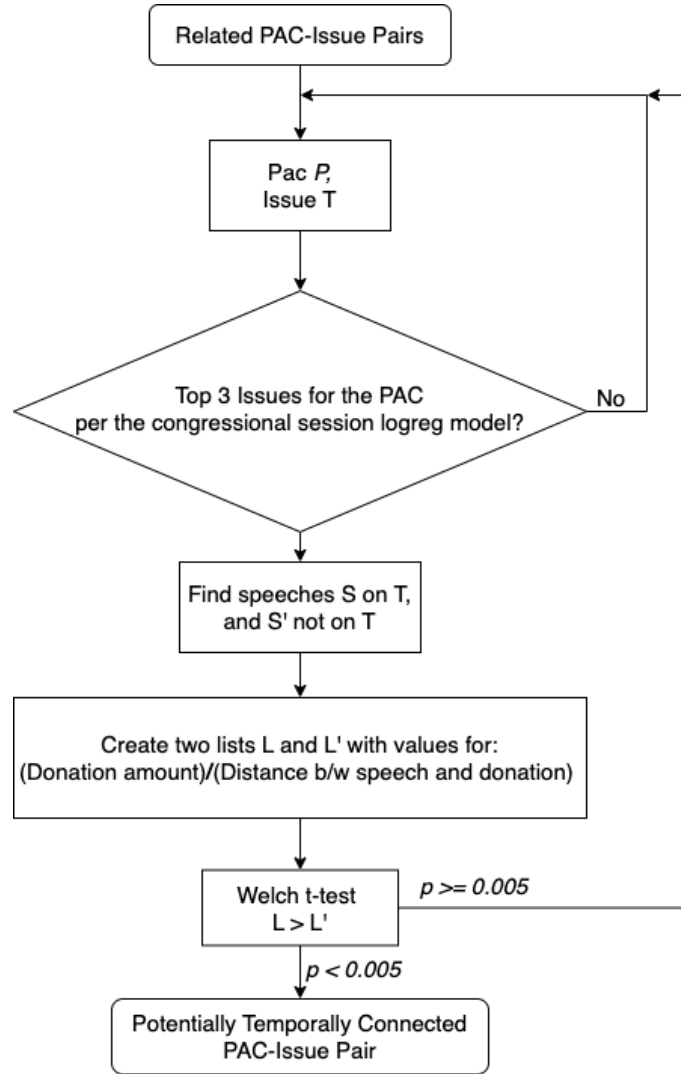


Figure 3.7: Algorithmic workflow of identifying relevant PAC-issue associations showing cases of significantly connected speech and donation events — cases where within a short time window (proximal), a donation made by a PAC interested in a particular issue and a speech made by the recipient on that issue co-occur.

terms of expert judgment.

3.7 Procedure for Surfacing Potential Temporal Connections

Our framework, including topic modeling and multinomial logistic regression, was both used on the entire data used all at once and also used for each individual

congressional cycle separately. Since we want to uncover potential connections between donation and speech events that occur in proximity to one another, we use the topics uncovered in a particular cycle in order to use the more fine-grained modeling as the starting point.

For each congressional cycle, we use that cycle’s regression model weights (β_1 in Eq. (3.1)) to identify the top topics for a PAC. We match a particular cycle’s topic modeling outputs with the topic modeling outputs of the model run on the entire data of floor speeches by using the top terms in the topic-word distributions: if at least five out of the top twenty terms for the topics are the same, they are considered a match. Future work should ideally obtain labels and PAC-issue relevance judgments for every congressional cycle, but since that would entail the annotation procedure being carried out twelve more times, we use our topic-matching proxy in order to enable the use of the results of human annotator ratings for PAC-issue associations made using the regression model outputs trained on the entire data.

A topic (provided by the topic model trained on a particular congressional cycle) is considered *relevant* for the PAC if the corresponding issue (uncovered by the topic model trained on the entire dataset) is *a)* among the top 10 topics for the PAC per the logistic regression model trained on the entire dataset, and *b)* rated as relevant (3 on the 3-point Likert scale) by at least one of the two independent experts. These relevant PAC-topic pairs for every cycle are then used as the starting points for our method outlined below as an algorithmic workflow. Our algorithm or procedure is also visualized as a flowchart in Fig. 3.7.

Our aim here is to filter the aforementioned relevant PAC-topic (or PAC-issue) pairs to those that show a potential temporal connection, *i.e.*, the PAC is not just associated with the policy issue, but also seems to donate closer to the date a speech

on that issue occurs. To do so, we iterate over each topic t and PAC p in each of the relevant topic-PAC pairs, and:

1. We use the document-topic posterior learned by the topic model (trained on the particular cycle's floor speeches) to identify speeches that are 'on' the topic t (based on the topic proportion in the speech being more than uniform and the topic being in the top 3 for the speech) and also the speeches in that cycle that are not 'on' that topic t (based on the topic proportion being less than uniform and the topic ranked in the bottom half of all topics for that speech).
2. For each of these two lists of speeches (on the topic t , and not on the topic t), we take the date of each speech, and find the closest date of donation made by the PAC p to the legislator who gave that particular speech. We record that (donation amount) divided by the absolute distance between the date of that speech and the date of donation to the legislator giving that speech by PAC p – the idea being that a high donation amount by PAC p given to the legislator close to their speech on topic t is what signifies a potential temporal connection.
3. We thus have a list of values for speeches on topic t and a list of values for speeches not on that topic. We conduct a Welch t-test, which does not assume equal population variance ([Welch, 1947](#)), to identify cases where the first list represents a distribution of values that is significantly greater than the second list ($p < 0.005$).
4. Based on the above significance test, we select relevant PAC-topic pairs with a potential temporal connection for the cycle. We can then look at speeches

made on the topic and cases of the relevant PAC making a donation much higher than their mean donation amount during that congressional cycle to the legislator who gave that speech within a particular time window around the speech. Some examples of such speeches and donations are shown in the Table [B.6](#).

5. Based on aggregating cases of such speeches, we can look at PAC and legislators that recur in such temporal connections, i.e., the PAC donating a larger-than-their-usual amount to a legislator giving the speech on a relevant issue, in, say, a 7-day period around the date of that speech, and see if that same PAC-legislator pair occurs multiple times either during a cycle or even across cycles. One illustrative example of such cases uncovered by our method is shown in Fig. [3.8](#); more such illustrative examples of automatically discovered PAC-legislator relationships in terms of patterns of donations in near-vicinity of topically relevant speeches made are included in Appendix [B.5](#) (Figs. [B.11](#) and [B.12](#)).

3.8 Discussion and Conclusion

Studying the connection between PAC donations to elected representatives and the issues those representatives prioritize in their speeches on the floor of Congress sheds new light on two important questions: *a)* How do organized donations affect the behavior of elected representatives? and *b)* Why do elected representatives prioritize one issue over another in their rhetoric? Floor speeches are an underexplored facet of the first question, while PAC donations are an underexamined factor in studies of issue salience or *attention* in Congress derived from floor speeches.

WASHINGTON

For-profit colleges giving big to helpful House members

Fredreka Schouten and Christopher Schnaars USA TODAY

Published 7:50 p.m. ET July 23, 2013 | Updated 11:26 a.m. ET July 24, 2013



WASHINGTON — House Education Committee Chairman Rep. John Kline, who saw a dramatic upsurge in campaign contributions from for-profit colleges in recent months, is pushing legislation that would help the industry preserve its access to federal student loans.

Rep. John Kline is chairman of the House Education and the Workforce Committee. Charles Dharapak, AP

Career Education Colleges and Universities

From Wikipedia, the free encyclopedia

Career Education Colleges and Universities (CECU) is a Washington, D.C.-based trade association that represents for-profit colleges.

Figure 3.8: An example of the kind of real-world relationships that the outputs of our modeling and analysis can capture automatically — for a PAC (*Career Education Colleges & Universities*) and an issue (*Education*) that our model (as well as experts) deem related to one another, we find that in a particular congressional cycle (2015-16), this PAC is significantly likely to donate a more-than-their-usual amount of money closer to dates on which speeches prioritizing the particular issue of education were made in the 2015-16 cycle. We found that such a significant connection is primarily driven by donations made to a particular legislator: *John Kline (R-MN)*. This legislator made a number of speeches relating to the issue of education on Feb 25, 2015 and Feb 26, 2015 – manually looking at the speeches, the legislator is arguing against federal funding for education and schooling and in favor of charter schools. The PAC made a donation of \$5000 to this legislator on Feb 27, 2015 – when the PAC’s average (std dev) donation amount to *any* legislator in that particular cycle was $\$1862.16(\pm\$1288.27)$. These kinds of relationships can be uncovered by journalists, but it can be a time-intensive process. Our data plus modeling can find such relationships (by ranking them towards the top) without supervision, and therefore, our work can enable journalists to focus on cases deemed significant by our analyses as starting points for further investigation. Snapshots in the image come from <https://www.usatoday.com/story/news/politics/2013/07/23/for-profit-colleges-contributions-house-regulations/2579041/> and the Wikipedia article for Career Education Colleges and Universities.

This research expands the ways in which we think about the potential impact of donations on policymaking and political representation.

Our key finding is that information about which PACs did and did not donate to a US House representative is meaningfully associated with representatives' issue-attention derived from their floor speeches. Recognizing and quantifying this association is important for understanding the dynamics of political power as exercised through agenda-setting and control in Congress. We find that donors (PACs) are significantly more associated with legislators' issue-attention than other conventional factors (such as committee assignments or partisanship). This association is made clearer in two different ways: first, the relationship holds for each congressional cycle over a period of twenty-four years, demonstrating that our findings are based in a consistent pattern for the US House of Representatives; and second, our predictive machine learning model can automatically learn meaningful issue-PAC associations, highlighting the strong patterns of PACs donating to those legislators who prioritize issues relevant to the PACs' own policy area of operation.

Our work focuses on analyzing the association between issue-attention and explanatory variables such as PACs and does not make claims about causality.¹⁹ We cannot say whether the relationship reflects donors rewarding legislators for their support on the House floor, whether donations spur a legislator to talk about an issue more than they would have otherwise, or whether both of these dynamics are happening. However, what our study uniquely shows is that there is a clear positive relationship between who gives money to politicians and what those elected officials talk about. Establishing this positive relationship between money and speech is a critical first step toward a better understanding of the important relationship between donors and legislators. With our novel focus on issue priorities expressed

¹⁹The association here is studied in terms of predictive capacity on held-out data.

in language (as opposed to roll-call votes), combined with our methodology and the new data presented, future research is better positioned to examine the age-old question of causality with new tools. Future work should explore legislators who ‘break character’ by talking about an issue they generally do not pay attention to in a particular period of time, and systematically analyze if donors interested in that issue tend to donate to the legislator before or after those unexpected speeches. This could imply an attempt to influence the legislator, or a reward for the difference in activity, respectively.

We acknowledge the possibility of other explanatory variables not considered in this study that could potentially predict legislators’ issue-attention better than information about their donors. However, since donor information offers significantly more predictive power than standard explanations, does so consistently over time, *and* constitutes a strong enough pattern that the automatic discovery of meaningful issue-PAC associations is possible, our results present a useful finding and a foundation on which future work can build. Future studies can consider other variables or attributes such as legislators’ race and gender, or attributes of their districts such as whether the district is closely contested across different parties in local and national elections.

The release of our new database, our codebase, annotations of topic modeling output (in order to obtain issue labels), ratings for issue-PAC associations, and other procedures and modeling outputs are intended to help spur future research into both the impact of donations on congressional activity as well as possible factors explaining agenda setting on the floor of Congress. One particular avenue for future work is to consider the various aspects and stances within a political issue present in floor speeches given by a legislator and move beyond *issue-attention* to *framing*. Now that we know that donors are meaningfully associated with how much legislators

talk about an issue, the next question can be: are they also associated with legislators' stances and framing choices in the context of a particular issue?

Another avenue for future research is to consider a different aspect of agenda-setting: what is *not* being said, as opposed to what is being said. From a broader theoretical perspective, the issue of non-events (such as what is not said) is an important consideration rooted in classic work on power and politics (Dahl, 1956; Bachrach and Baratz, 1962, 1963; Gaventa, 1982). Our current methodology does not support modeling legislators who do not engage in any floor speech activity. Our dataset, however, can help drive research with models that do not have this limitation that can effectively study those legislators who do not give floor speeches. Such future work can better understand the impact of various sources of information about legislators, such as their donors, on this other aspect of agenda-setting. We note that our current framework *does* model the complete topic distribution or issue-attention, and therefore considers legislators not speaking about some issues at all while they prioritize other issues. Hence, what is not being said on specific political issues is a part of our modeling and drives our findings; we leave the specific analysis of such issue-specific gaps in legislators' floor speech activity to future work as well.

Apart from scholarly work, our data and identification of issues and PACs annotated as related to those issues can serve as the core of an *enabling technology*. It can help journalists identify donation and speech events in Congress that deserve scrutiny. Specifically, we contribute a procedure that uses our trained machine learning model outputs and expert judgments of issue-donor associations to help find significantly-connected speech and donation events occurring in close proximity to one another. Identifying such pairs of events can help uncover significant PAC-legislator relationships, such as the example provided in Fig. 3.8 — easily

validated by using a search query of the form '[PAC Name] [legislator name]' in a search engine. Examples of connected speech and donation events along with more examples of PAC-legislator relationships discovered using our approach are provided in Appendix [B.5](#), and demonstrate the potential for increased public transparency beyond what raw donation transaction data or anecdotal evidence alone provides.

Chapter 4: Context-based Variation in Communicative Choices: US Legislators' Ideological Expression Across Audiences

Members of the US Congress face both consistent pressure and ample opportunity to express their ideological positions. The most commonly measured outlet is congressional floor votes, but it is not the only one. Congress members also regularly express themselves in their rhetoric, which includes speeches on the floor of Congress, press releases and interviews with media, their tweets and other forms of social media content, and more; in fact, these forms of communication can offer a more nuanced analysis of ideology and beliefs ([Diermeier et al., 2012](#); [Grimmer, 2013](#); [Grimmer and Stewart, 2013](#)).

In particular, members have opportunities to speak on the chamber floor, where their policy discussions become part of the congressional record, and Twitter provides legislators with an opportunity to communicate their beliefs briefly and directly. Using large-scale textual records of these communiques, unsupervised NLP methods can exploit the patterns of communicative choices made by these political actors to measure ideological positioning and help enable rich and substantive exploration of *variations* across different contexts. This is valuable since politicians navigate the often interacting constraints of different audiences, communication styles, policy issues, motivations, and their own underlying ideology.

In this chapter, we use an existing unsupervised NLP tool — *text-based ideal points*

(henceforth referred to as TBIP, [Vafa et al., 2020](#)) — to investigate the understudied connection between ideology, communication style, and audience. In particular, we examine how members of the US House of Representatives express themselves differently across their votes, their floor speeches, and their tweets. TBIP can derive scalar estimates of ideological positioning or ideal point estimates for authors present in a dataset using their texts alone. We establish the stability (automatically) and validity (through extensive human assessments) of the outputs of the method, and enable the method to use a different topic model such as LDA for its initial topic estimates (as opposed to the less stable Poisson factorization based estimation the method originally relies on). We construct a new dataset combining different kinds of information about the representatives, and contribute new understanding of legislative behavior and ideological positioning of members of the US Congress.¹

In the previous chapters of this dissertation (Chapters [2](#) and [3](#)), we regarded topics discovered by a topic model as the operating communicative choice. However, each categorical topic or issue itself can span a range of discussions — for example, certain aspects can be emphasized or selected in order to frame the same issue differently, and these also constitute *communicative choices*. This view mirrors an understanding of *framing* as *second-level agenda-setting*, where first-level agenda

¹The contents of this chapter are adapted from joint work that is under review as: SoRelle W. Gaynor, Kristina Miler, **Pranav Goel**, Alexander M. Hoyle, and Philip Resnik. 2023. “Do You Walk the Walk, Talk the Talk, or Tweet the Tweet? Legislators’ Ideal Points Across Venues”. Sources of the adaptation here also include the manuscript presented at *Women in Legislative Studies Research Seminar, Fall 2022*: SoRelle W. Gaynor, Kristina Miler, **Pranav Goel**, Alexander M. Hoyle, and Philip Resnik. “Do You Walk the Walk, Talk the Talk, or tweet the tweet? Legislators’ Ideal Points Across Venues”; and the manuscript presented at *Text as Data, 2021*: **Pranav Goel**, SoRelle W. Gaynor, Alexander M. Hoyle, Kristina Miler, and Philip Resnik. “What Text-Based Ideal Points Reveal about Strategic Considerations behind Congressional Behavior across Different Outlets and Issues”. Goel led the work on modeling as well as compiling various legislator data with their ideal point model in order to enable all the analyses that constitute this work.

setting can be seen as selecting the objects for attention (or ‘what’ to think about) and second-level agenda setting can be seen as selecting the attributes for ‘how’ to think about the object (McCombs, 1994; McCombs and Ghanem, 2001).² It is these within-issue framing choices cast as ‘votes’ on a topic or issue (the ‘bill’) that form the patterns used by TBIP in order to estimate ideal points for members of Congress or legislators (as the authors of the floor speeches or tweets; more details in Section 4.5). The model discovers ideological frames deployed within each discovered topic and estimates one scalar ideal point value for each author (Congress member).

The new dataset we put together to enable this work including model outputs, human annotation instructions and results to establish model viability, and our entire codebase for applying the model and obtaining the outputs at the heart of our analyses is publicly available at <https://zenodo.org/record/7641732> (DOI: 10.5281/zenodo.7641732).

Roadmap for this chapter. We will first introduce and summarize the contributions of the work presented in this chapter (Section 4.1) and provide a comprehensive overview of relevant prior work as background (Section 4.2). We will then make the theoretical case for having multiple measures of ideological positioning (Section 4.3). Next, we will discuss the data we use in this study (Section 4.4) followed by an exposition of the method (Section 4.5). We will discuss results and different analyses examining differences in ideological positioning across different venues (Section 4.6), and conclude by summarizing our findings and discussing their implications (Section 4.7). We will also provide insights into the limitations and important caveats of this work (Section 4.8).

²Various computational methods, such as Nguyen et al. (2015b), have proposed hierarchical models motivated by this view of agenda setting.

4.1 Introduction

As we noted above, members of the US Congress simultaneously face different kinds of pressures and are afforded ample opportunity to express their ideological positions. While congressional floor votes (on bills and motions introduced in Congress) are traditionally the most commonly measured outlet, there are plenty of other outlets of import, including ones that make use of spoken or written language. We consider two such rhetorical outlets in this work: floor speeches and Twitter tweets. Each of these three actions — voting, floor speeches, and tweets — allows members to express their ideological positions to different audiences. Yet, in part due to measurement challenges, scholars lack an understanding of how members' ideological expressions differ by medium or venue.

By making use of an available tool in TBIP ([Vafa et al., 2020](#)), we seek to help fill this gap and derive three measures of ideology for representatives in the US House across two congressional sessions (2015-16 and 2017-18). These three measures allow us to examine the differences across congressional ideal points in the aggregate and provide novel evidence of how individual legislators position themselves differently depending on the venue. This comparison enables an evaluation of an understudied interaction between ideology, communication style, and audience. Ultimately, we find that members express different ideological positions across these different venues, and we argue this is due to variation in the audience.

Text-based measures in particular allow us to capture ideological nuance that allows legislators to differentiate themselves from their peers in ways that voting does not allow. Voting is constrained by both the dichotomous nature of the “yea or nay” structure inherent to voting, as well as the strong partisan divide that exists

in the modern Congress.³ As a result, ideal points derived from roll-call votes provide legislators with a more limited expression of their ideological positioning. Conversely, floor speeches and social media provide members an opportunity to explain their legislative decisions and allow them to take more moderate or more polarized positions.

Furthermore, across these three venues, members of Congress knowingly contend with different audience considerations when making ideological decisions or communicating their position. We evaluate constituent effects, including the partisanship, demographics, and policy interests of their district, as well as institutional goals such as party status or personal responsibilities of their position in their chamber, whether in leadership, committee, or caucuses, on an individual legislator's ideological positioning. Lastly, we consider individual demographic attributes of the legislator, such as gender and race.

Our work engages the considerable literature on the measurement of ideology and congressional representation, paying particular attention to the demands of district and chamber responsibilities. Utilizing an unsupervised NLP method (after validating its outputs, and making both data processing-based and methodological adjustments), this work helps improve the theoretical understanding of the motivations behind legislative behavior; specifically, the ways in which ideological positions differ by forum or venue.

We use TBIP to develop three reliable estimates of ideology for members of the US Congress: a vote-based ideal point and two text-based ideal points based respectively on congressional speech and official tweets. By ranking members across these three measures, we observe if, and how, members shift their ideological

³This was a salient motivation for our work in the previous chapter examining the association between donation activity and congressional rhetoric; Chapter 3.

position across these three mediums and audiences. We apply these measures to an original dataset of members of the 115th (2015-16) and 116th (2017-18) Congress that includes demographic, institutional, and district information.

We find that different means of legislative expression do produce different ideal points, both at the aggregate and individual levels. The distribution of House members' ideal points differs considerably when measured by votes, speech, and tweets. Across our dataset, floor speeches and tweets allow legislators to differentiate themselves from their co-partisans and to take more nuanced positions. In particular, we find that ideal points expressed via social media offer legislators the opportunity to express more ideologically extreme positions as compared to vote-based ideal points.

Interestingly, these effects differ by gender and race. We also find notable differences between the two major political parties that are reflective of the demographic and ideological challenges facing the two caucuses. The Republican caucus is more united in both vote and speech, compared to the Democratic caucus. Overall, we argue that the ability of legislators to adjust their relative ideological position across venues is valuable because it allows them to better represent their districts. Thus, using multiple measures of ideology — directly enabled by considering and using communicative choices made in textual data — has implications for assessing ideological congruence and representation in Congress.

4.2 Background

There is a large congressional literature on the meaning, role, and measurement of ideology (*e.g.*, [Jackson and Kingdon, 1992](#); [Krehbiel, 1993](#); [Hill, 2001](#); [Bishin, 2003](#); [Clinton et al., 2012](#)). Legislators' ideology is generally described as their

position along the liberal-conservative spectrum that reflects the policies they would implement if they could (*e.g.*, [Poole and Rosenthal, 1985, 1991](#); [Clinton et al., 2004](#)). Put differently, legislators' ideology is seen as their expressed policy preferences and measured by observable actions.

The primary audiences for this ideological positioning include legislators' constituents and congressional colleagues. Members of Congress are famously reelection-minded and use numerous opportunities to communicate their ideological positions to their constituents ([Mayhew, 1974](#)). For constituents, legislator-led communication serves as the connecting link between home and Washington, D. C. Given the wide range of policymaking activity, constituents are reliant on members of Congress to pass along relevant information to their districts and states. Even the most attentive constituent will find it challenging to keep up with every committee hearing, floor proceeding, or bill introduction ([Arnold, 1990](#)). Members utilize constituent communication to share information about legislative activities ([Hall, 1987](#); [Curry, 2015](#)), constituent services ([Grimmer, 2013](#); [Parker and Goodman, 2009](#)), partisan messages ([Jacobson and Carson, 2019](#)) and partisanship signaling ([Hemphill and Shapiro, 2019](#)), and how and why they voted ([Ansolabehere and Jones, 2010](#)). Increasingly, a large part of constituent communication focuses on members' ideological beliefs ([Cormack, 2016](#); [Russell, 2018](#); [Smith and Russell, 2022](#)). Members are eager to promote not only their legislative and partisan accomplishments, but to provide ideological reasoning for these decisions too.

Over time, the changes in communications and technology have created new opportunities for messaging. Constituents no longer rely on newspapers and nightly news for updates on their member of Congress. Anyone can go online and find the roll-call outcome of any vote, on any bill or amendment. The advent of C-SPAN and the near-constant television presence on the floor and in committees

have made floor speeches accessible to constituents even if not widely watched. Constituents can, and do, directly and immediately contact their members via social media and email. Over the past decade, social media has also become a prominent form of congressional communication: In 2020, every member of Congress had a dedicated Facebook page for communication, 99% ran an official Twitter account, and 86 percent of Senators (79% of House members) had an official Instagram account.⁴

Members also communicate with one another. Classic studies underscore that legislators look to their colleagues for cues when issues come before Congress, and are a factor in understanding legislative decisions (*e.g.*, [Kingdon, 1989](#); [Matthews and Stimson, 1975](#)). Actions taken inside Congress, therefore, also can be used to signal one's position to congressional peers. Votes serve as a signal to party leaders that have control over the agenda. Floor speeches and social media allow members to add more depth to their ideological expression. Research has found that some members use verbal communication, like floor speeches, as a way to signal their institutional status. [Lerner and Shaffer \(2020\)](#) found that the most prominent speakers in committees were also more effective lawmakers. Research on social media in Congress suggests that it too may play a role in intra-congressional relations (*e.g.*, [Barberá, 2015](#); [Barberá et al., 2019](#); [Shapiro et al., 2012](#)). Unlike votes or floor speeches, Twitter is an equal-opportunity communication source. All members have the ability to tweet any time of day, regardless of position or whether Congress is in session or not, and about (nearly) any topic.⁵ Thus, communication is used by lawmakers to send a signal back home or ideological disposition to congressional

⁴<https://www.quorum.us/reports/2020-congressional-social-media-report/>

⁵Official Twitter accounts are monitored by the House Committee on Ethics and House Communications Standards Commission and are subject to rules on campaign fundraising and personal information.

peers. Ideological positioning (and party alliance) can be demonstrated on social media, just as it can be through votes and floor speeches.

Existing measurements of ideology. Political science has tackled the measurement of ideology in several ways. The most common and widely accepted application is spatial models that rely on roll call votes (*e.g.*, [Poole and Rosenthal, 1991](#)). Roll call vote models have been applied to understand the ideology of U.S. Congress members, as well as state legislatures ([Shor and McCarty, 2011](#)). These models establish a continuum, from liberal to conservative, that allows researchers to orient individual legislators or groups (such as a state legislature) on an ideological spectrum. While there are limitations to this approach — most notably that the data is both binary (yes or no vote) and is limited to the legislation that is brought forward for a vote — these models are easily applicable and well-established. Other iterations of spatial models have provided insight beyond chamber dynamics, using external relationships such as campaign donations ([Bonica, 2013](#)) or interest group ratings ([Berry et al., 2010](#)).

As the accessibility of computational text analysis has increased, spoken and written words have been used to establish ideology, sometimes in tandem with legislative votes. Text allows researchers to consider policy positions beyond floor votes, and contributes to a better understanding of ideology, particularly for policies that are not subject to roll call votes. Spatial models using texts have been applied to the U. S. Congress to provide a more nuanced understanding of legislators' positioning, including expressed preferences by policy area (*e.g.*, [Diermeier et al., 2012](#); [Gerrish and Blei, 2011](#); [Grimmer, 2013](#); [Lauderdale and Clark, 2014](#); [Kim et al., 2018](#); [Vafa et al., 2020](#); [Ebanks et al., 2021](#)). Language has also been used in conjunction with vote behavior to give a more holistic view of legislative dynamics.

For example, [Nguyen et al. \(2015b\)](#) develop a model of legislative votes, speech, and bill text to examine substantive policy differences between the Tea Party caucus and establishment Republicans. Relatedly, [Davoodi et al. \(2020\)](#) use relationships between bill text, donor contributions, and legislator attributes to understand the factors that characterize (dis)agreement within state legislatures.

Text-based analysis also has proven especially valuable in establishing ideology in institutions such as the judicial branch, in which quantifiable measures such as votes or campaign donations are limited (*e.g.*, [Bonica and Sen, 2021](#); [Hausladen et al., 2020](#); [Lauderdale and Clark, 2014](#)), as well as electoral contexts such as debate reactions ([Argyle et al., 2021](#)). Text analysis has also been used to assess the ideological positions of executive branch officials ([Bertelli and Grose, 2011](#)) and presidents ([Treier, 2010](#)). The application of text analysis-based ideal points is not limited to the U.S. context as scholars have examined party platforms and legislative speeches in numerous countries (*e.g.*, [Laver et al., 2003](#); [Benoit and Laver, 2012](#); [Lowe et al., 2011](#)).

Ideology and representation. Scholars' enduring interest in legislators' ideological positioning in part reflects the importance of questions about congressional representation, and the promise of ideal points to measure the extent to which members of Congress accurately reflect the preferences of their constituents ([Miller and Stokes, 1963](#)). Consequently, there is a sizeable literature focused on ideological congruence, or whether ideologically more liberal (conservative) districts are represented by more liberal (conservative) legislators (*e.g.*, [Ansolabehere and Jones, 2010](#); [Carson et al., 2010](#); [Hall, 2015](#); [Miler, 2018](#)). The increase in polarization in American politics raises questions of whether members of Congress are more ideologically extreme than their voters, or whether legislators are responding to the preferences

of their constituents (see, [Tausanovitch and Warshaw, 2013](#); [Barberá et al., 2019](#); [Battista et al., 2021](#)). There is some evidence that voters punish legislators whose preferences are out of line with the preferences of the district ([Ansolabehere et al., 2001](#); [Canes-Wrone et al., 2002](#)), but the electoral penalty for being out of step may be declining ([Bonica et al., 2018](#); [Highton, 2019](#)), and there is evidence that legislators are more extreme than the public ([Bafumi and Herron, 2010](#)).

The lack of a clear relationship between legislator ideological positioning and public preferences may reflect the fact that studies of ideological representation generally look at vote-based ideal points, and voting may not be the behavior we expect to be most reflective of constituents. In part, the scarcity of constituent ideal points that are comparable to a vote-based ideal point hampers the study of ideological congruence. For instance, whereas a legislator votes on a specific piece of environmental legislation that will change the accepted levels of a specific pollutant, public opinion surveys generally ask broad questions like whether the government should prioritize doing more on the environment.

In addition, efforts to measure ideological representation will falter when considering issues where the public and legislators have preferences, but those issues do not receive a roll-call vote. Thus, the expanded breadth of measurement of legislative ideal points derived using TBIP in this work can also facilitate a re-examination of ideological representation by engaging the question of which ideological expression would be expected to reflect constituents. This is due to the fact that texts including transcripts on the floor of Congress as well as social media posts can be very public-facing and are not bound by the same constraints as votes. We argue that text-based ideological expressions are taken with an eye to an audience of constituents and therefore offer a valuable tool for better understanding congressional representation.

4.3 Theoretical Motivation: The Case for Multiple Measures of Ideology

Legislators make choices about how to present themselves and communicate their actions to wider audiences (*e.g.*, [Fenno, 2003](#); [Mayhew, 1974](#); [Grimmer, 2013](#); [Blum et al., 2022](#)), and we expect that this behavior extends to how they express their ideological positions. We use TBIP ([Vafa et al., 2020](#)) to develop comparable ideal point models for legislators across three forms of behavior: votes, legislative speech, and social media expression. Our research design then examines the motivations behind each type of ideal point as well as the dynamics of shifts in an individual legislator’s ideal points across venues. While votes provide the most limited choice and ideological expression, they can present a clear signal of support for the party. Voting is primarily an inside-Congress audience, including party leaders and colleagues. And while larger or controversial legislation does attract the attention of attentive voters, these are often highly partisan constituents. Floor speeches allow for greater ideological variation, but still face time and scheduling restrictions subject to party leaders and member travel.⁶ The audience is limited to those in the chamber (or watching C-SPAN), unless the member proactively promotes their floor speech through another medium, or a third party publicizes the speech more broadly. Social media has the fewest limitations in terms of content and timing, but a short word count limits the actual text a member can share, particularly on Twitter. Notably, Twitter provides individual members with equal footing with party leaders, allowing members to speak directly to a national audience. Some rank-and-file members have easily surpassed the attention and audience of the

⁶Party leaders control the floor schedule, including when members are able to speak on the floor. Furthermore, in the House, the majority of speeches are limited to one minute.

formal leaders of their party.

We expect that members will take audience variation into consideration, and present different ideological stances across these three mediums.⁷ Given that floor votes are determined (and inherently limited) by party leaders, we expect that vote-based ideal points will be the most polarized ideal point, shaped mainly by party and chamber dynamics. We expect floor speech-based ideal points to be shaped by a combination of both institutional and district motivations, thus introducing more variation across legislators' ideal points. For example, we consider how party and chamber responsibilities, like seniority and leadership positions, may lead some members to reflect a more unified, party-centered message. However, given the public-facing nature of floor speeches (and the tendency to use them in press releases) floor speeches also allow members to appeal to their district's nuanced ideology — whether that be more moderate or more extreme. Conversely, because tweets carry little institutional weight, we expect tweet-based ideal points to reflect a legislator's positioning for a more national constituency. Given previous scholarship on extremism on social media (*e.g.*, [Banks et al., 2021](#); [Conover et al., 2011](#); [Barberá, 2015](#); [Ballard et al., 2023](#)), we anticipate greater variation in social media posts: some legislators' tweet-based ideal points are likely to be more ideologically extreme, while other legislators may deliberately take more moderate positions.

These expectations not only shape how we expect the three ideal point measures to compare with one another but also how we expect individual legislators to use these distinct opportunities to communicate their ideological position. We predict that a member will shift their ideological position as expressed through speeches on the House floor or on Twitter (as compared to roll-call votes) to appeal

⁷We note that prior work in [Jost and Sterling \(2020\)](#) demonstrates that even when expressing similar kind of ideological stances, legislators may emphasize different values across different platforms — Twitter, Facebook, and the floor of Congress.

to constituent and partisan audiences. However, we expect institutional factors – such as party leadership and seniority – to impact ideological positioning, as well. These party-loyal members will be less likely to vary from their vote-based position.

4.4 Data

To estimate different ideal point measures using TBIP, we utilize votes, congressional records, and tweets from the official accounts of House members from the 115th and 116th Congress (2017-2020). We begin by creating an original dataset with these three TBIP measures for each member alongside biographical and district information for all members of the 115th and 116th Congress (the US House of Representatives). Thus, unlike prior research, we include information from two Congresses, across all policy issues. The result is a dataset with 507 members.⁸

For each member, we collected biographical information such as gender, race, party, the number of terms they have served in the House, whether they are in a position of party or committee leadership or on a high-ranking committee (Appropriations, Energy and Commerce, and Ways and Means), and if they are a member of a major intra-party caucus. This includes the Problem Solvers Caucus, Republican Study Committee (RSC), Freedom Caucus, Blue Dog Democratic Caucus, New Democratic Coalition, and Progressive Caucus. Caucus information was collected via archival research using the Congressional Yellow Books ([Gaynor, 2021](#)). While most representatives' membership spans across both congressional sessions, legislators are coded as being a member of the caucus if they were a member at any time in the dataset, *i.e.*, the 115th or the 116th Congress.

⁸The majority of representatives are members of both the 115th and 116th Congress. This dataset also includes those members who departed from the 115th Congress or were new members in the 116th Congress.

We also collect district information, including the percentage of minorities in the district, the district unemployment rate, a categorical variable for district density that indicates the urban, suburban, or rural nature of the district, and electoral information on both the House member (percent of the vote received in the most recent general election) and the 2016 Presidential election (percent of the vote received by the Republican presidential candidate), which we consider to be a proxy for citizens' national political preferences (Tausanovitch and Warshaw, 2013).⁹ These legislator-specific sources of information serve as our independent variables in our pursuit of understanding the factors that help explain the variation in ideological communication strategies across different venue or audience-based contexts.

To acquire the text data we use to apply the TBIP model, we use the Congressional Record parser made available by Judd et al. (2017) to download floor speeches and extract the speech text and speaker Bioguide ID (to correlate with biographical information of lawmakers).¹⁰ For tweet data, we acquire the Twitter user IDs for House representatives and download the accounts' text using Python package TWEETPY (Roesslein, 2020).¹¹ We scrape roll-call vote data using the public domain data collectors for US Congress, primarily maintained by the Sunlight Foundation and GovTrack.¹²

⁹District demographic information is collected using the Census Bureau's American Community Survey data, 2016. For district density information, we use a simplified measure of the CityLab Congressional Density Index. The categorical mapping we use for the simplification: Rural = pure rural, rural-suburban mix; Suburban = sparse suburban, dense suburban; Urban = urban-suburban mix, pure urban.

¹⁰<https://github.com/unitedstates/congressional-record>

¹¹For Twitter user IDs for House representatives — 115th Congress: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/UIVHQR>; 116th Congress: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/MB0JNS>.

¹²<https://github.com/unitedstates/congress>

Preparing text data. Using the text processing settings described in [Vafa et al. \(2020\)](#) as our guide, we removed legislators with a low number of speeches or tweets. From the text itself, we also remove the names of cities, states, and representatives, as well as common English stopwords and corpus-specific stopwords such as procedural terms that are unlikely to be relevant or useful in distinguishing documents and speakers. To reduce noise, we omit terms spoken by only a handful of representatives. Lastly, we remove short procedural speeches such as ‘I yield back the balance of my time’, which are frequent occurrences that offer no ideological or policymaking value, following the process provided in [Card et al. \(2022\)](#).

For tweets, we engaged in similar standard preprocessing for text as specified above in terms of stopword removal. Following [Vafa et al. \(2020\)](#), we retain hashtags. The original number of downloaded non-empty tweets was about 2.67 million, which was not computationally tractable for training the TBIP model, and therefore, we randomly sampled about 300,000 tweets, maintaining the original distribution of the number of tweets per author. Additional specific details and other major processing decisions for both datasets along with the final vocabulary size and number of documents used as input to the method are provided in Appendix [C.1](#).

4.5 Method

As noted above, we use TBIP to create three ideal point measures in order to compare ideology across votes, speech, and tweets. In this section, we directly use the exposition in [Vafa et al. \(2020\)](#) to provide a background on the TBIP method. We then successfully extend this method in order to facilitate the use of LDA topic modeling output as initial topic estimates — this constitutes the method used to perform our analyses in Section [4.6](#). The TBIP model requires only the collection of

text documents and who the author is for each document — it does not need party or any other labels for either the documents or the authors. The unsupervised setup should allow for easy adoption on different datasets and contexts, and crucially, our efforts establish the *stability* and *validity* of the model outputs that have not been established previously (these efforts are documented in Appendices C.2 and C.3).

In general, Bayesian ideal point models consider legislators voting yea or nay on a shared set of bills and posit that the vote v_{ij} by legislator i on a bill j can be modeled in terms of a per-legislator scalar latent variable (x_i) and per-bill scalar latent variables (α_j, η_j) as:

$$v_{ij} \sim f(\alpha_j + x_i \eta_j) \quad (4.1)$$

where x_i is the ideal point of the legislator indicating their polarity on a political spectrum, while η_j is the polarity of the bill.¹³

The resulting output is based on the following intuition: when x_i and η_j share the same sign, we can assume that the lawmakers are more likely to vote in a similar way. When lawmakers differ in their spatial direction, they are less likely to vote together.¹⁴

In brief, text-based ideal points (TBIP) are a generative model of text (represented by bag-of-word counts y_{dv}). Specifically, TBIP regards documents d (floor speeches, tweets, or any pieces of text) as an admixture of k latent topics (each represented

¹³In the general case, ideal points are not obligated to reflect individuals’ ideology. They may just as well correspond to demographic or regional differences, if these are the most important dimension of variation. Even in cases where the most obvious interpretation of the axis is ideological, it may in fact reflect party unity. Nonetheless, they remain useful in practice as likely indicators of politicians’ ideological stances.

¹⁴From Vafa et al. (2020) explaining the standard spatial roll-call vote-based models: “The per-bill intercept term α_j is called the bill popularity. It captures that some bills are uncontroversial, where all lawmakers are likely to vote for them (or against them) regardless of their ideology.”

by a vector of per-document topic intensities θ_{dk} .¹⁵ Topics are a distribution over the vocabulary V , the set of words used in the document collection (a vector of per-topic word probabilities $\mathbf{B}_{kv}$). θ and $\mathbf{B}$ are initialized from a factorization of the observed document-word co-occurrence matrix Y , composed of elements y_{dv} , denoting the count of word v in document d .¹⁶

Considering each $\mathbf{B}$ as a *neutral topic*, where individual word probabilities $\mathbf{B}_k$ are adjusted by author- and topic-specific latent variables according to the word's polarity. Two latent variables capture the polarity of the authors (lawmakers) and topics. A scalar ideal point x_{a_d} for the author a of a document d conveys the author's polarity. A vector η_k "polarizes" the topic by modifying its word probabilities.

$$y_{dv} \sim f\left(\sum_k \theta_{dk} \exp(\log \beta_{kv} + x_{a_d} \eta_{kv})\right) \quad (4.2)$$

We note that Vafa et al. (2020) use the Poisson distribution for f , and use this formulation to draw a parallel with the Bayesian vote-based ideal point estimation discussed above, with the $\log \beta_{kv}$ term being the intercept for the "popularity" of the word v in topic k , while η_{kv} represents the "polarity" of the word v in topic k . θ_{dk} ensures that it is the topics discussed in the document d that are contributing to the sum that estimates word counts for that document. η_{kv} being zero would mean the word is not politicized in that particular topic. In case of a non-zero η_{kv} , sharing the same sign as with the author's ideal point x_{a_d} would mean increasing the (Poisson rate of) word count for v in the document d , and decreasing the (Poisson rate of) count if the signs are opposite.

Thus, intuitively, the model evaluates the selection of words by legislators

¹⁵Here, topics refer to latent human-interpretable themes identified in a collection of documents by a probabilistic topic model (Hofmann, 2001; Blei et al., 2003; Blei, 2012), exactly as used in the previous chapters of this dissertation.

¹⁶Although θ and $\mathbf{B}$ may be randomly initialized instead of being learned using the observed y_{dv} .

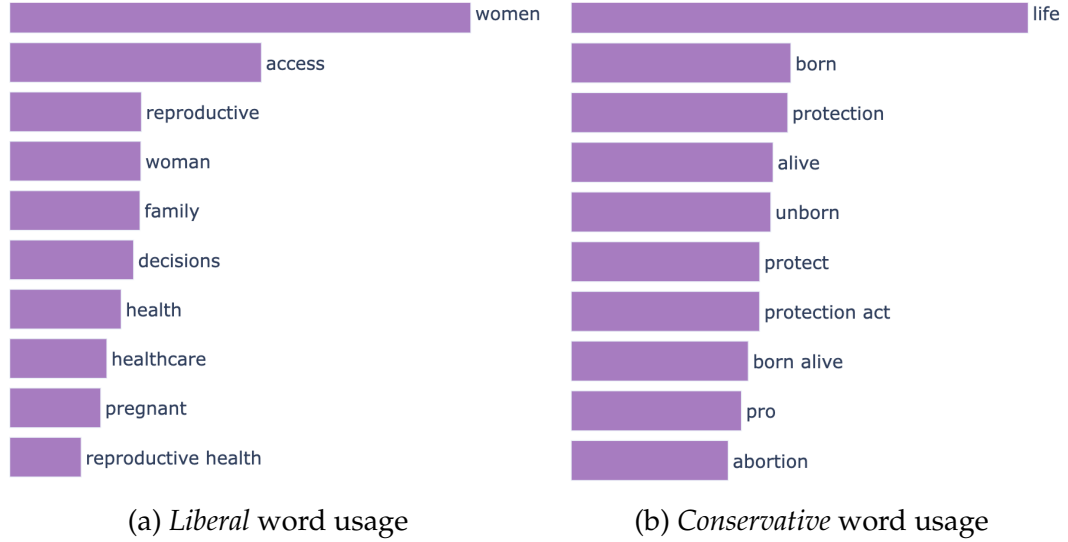


Figure 4.1: Illustration of “Liberal” and “Conservative” word usage on the issue of Abortion using our implementation of the TBIP method on our floor speeches data; generated using <https://github.com/keyonvafa/tbip>. The length of the bars indicates the strength (polarity) of the term v in topic k (η_{kv}).

on given “topics”, derived from texts authored by these legislators rather than votes on proposed congressional bills. During inference, latent variables capture the topic-specific polarity of individual words (η_{kv}). For example, for a topic concerning abortion, the words “women”, “unborn”, and “child” will likely have larger absolute values than for a topic about the environment. Conservative authors may use words like “life” or “unborn” with higher frequency than those who refer to abortion as a medical procedure (Fig. 4.1b). Conversely, liberal authors might use these words less frequently when discussing the same topic, instead using terms like “women” or “reproductive health” more (Fig. 4.1a). If we assume that estimation assigns conservatives a positive ideal point (> 0), and liberals a negative one (< 0), then “unborn” will have a positive sign and “women” a negative one. This neatly ties in with the concept of framing, which selects some particular perspectives within an issue to make more salient in discussions on that issue

(Entman, 1993). Framing has been shown to be an important tool of communication in politics, including influencing policy discussions and applications (e.g., Chong and Druckman, 2007; Nelson et al., 1997; Baumgartner et al., 2008).

Practical implementation details and extensions for analyses in our study. We find that the ideal point estimates for the authors derived by TBIP are highly stable, *i.e.*, with a particular set of initial topics, running the ideal point estimation step multiple times with different random initialization provides similar ideal point values for the authors in our dataset. This is an important concern for the reliable use of an unsupervised NLP method, allowing empirical research with a single model run with the assurance that the results and new theoretical understanding derived from those results will not change across different runs of the same model.¹⁷ Details are provided in Appendix C.2.

As explained previously, TBIP is a generative model of text, where individual word probabilities are adjusted by author- and topic-specific latent variables according to the word’s polarity. However, using a corpus of textual documents is very different from using voting records (Vafa et al., 2020). Votes have a cleaner structure: they are generally two-dimensional and associated with specific bills. Language data is high dimensional, unstructured, and ambiguous. The meaning and significance of a word are dependent on its context in multiple ways and may not even be relevant to political ideology.

The original TBIP model counteracts this challenge by running a generalized topic model step using Poisson Factorization. This provides us with the initial topic intensities in documents as well as neutral topics, which are then used in the ideal point estimation step. In order to obtain stable results, TBIP must be initialized

¹⁷Indeed, this is an important criterion in evaluating and comparing topic models in Hoyle et al. (2022).

with topic-word distribution $\mathbf{B}$ and document-topic distribution θ estimated from a separate topic model. In the standard setup, these are obtained through a Poisson factorization model. However, the outputs of models estimated with variational inference, like Poisson factorization, are less interpretable (Section 2.2) and stable (Hoyle et al., 2022) than those provided by LDA (Blei et al., 2003) estimated with Gibbs sampling (Griffiths, 2002).

To rectify this issue, we initialize with the outputs of such an LDA model using the MALLET (McCallum, 2002). However, $\mathbf{B}$ and θ are sampled from a Gamma distribution in Poisson factorization and Dirichlet distributions in MALLET LDA. To transform the posterior estimates of $\mathbf{B}$ and θ provided by the MALLET model, we use the following identity:

$$\begin{aligned} \text{for samples } y_1, y_2, \dots, y_n &\sim \text{Gamma}(\alpha, 1) \\ \text{and } [x_1, x_2, \dots, x_n] &\sim \text{Dirichlet}(\alpha), \\ \text{then } x_i &= \frac{y_i}{\sum_j y_j} \end{aligned} \tag{4.3}$$

To estimate the normalization factor ($\sum_j y_j$) for the document-topic and topic-word estimates, we run the Poisson factorization model 10 times and take the mean. We then scale the LDA estimates by the corresponding factor.

Our procedure makes the original TBIP procedure more flexible and adaptable by allowing the use of a different topic model of the practitioner’s choice to obtain initial topic estimates. An additional advantage of using MALLET’s LDA implementation to initialize topics is that the topics are generally relatively stable (an important consideration for content analysis and empirical research with text data) as established in Hoyle et al. (2022).

In addition, we utilized subject-matter experts to validate the coherence and polarization of the TBIP model’s outputs. Two political science graduate students independently first reviewed the topics produced by MALLET using our topic modeling curation protocol where top terms, as well as top documents for each topic, are provided.¹⁸ Both experts were asked if the topic represented an easily identifiable category or meaningful concept to obtain coherence ratings on a 3-point Likert scale, where 1 means ‘not identifiable’ and 3 means ‘easily identifiable’ topic. If the topic was rated 2 (somewhat identifiable) or 3, students then rated the topic’s *expected* polarization using their understanding of the policy issue represented by the topic and their knowledge of American politics. Ultimately, we found that our models performed very well: the overwhelming majority of topics were not only coherent, but the resulting polarized version of topics found by TBIP performed along ideological lines as expected — the topics were generally polarized (or not) as expected, and further, the direction of polarity of the frames found within each topic, *i.e.*, the version of topic identified as liberal or conservative by the TBIP model was also validated by our independent coders. This process was repeated for both the floor speeches and tweets; details of the annotation or human judgment process as well as results are provided in Appendix C.3.

Finally, we normalize the ideal points so that they can be compared directly. Vote-based, speech-based, and tweet-based ideal points are generated using independent models, and are not directly comparable — 0.1 of text-based ideal points are not “ideologically equivalent” to 0.1 vote-based ideal points. Since we are concerned with legislators’ *relative* positioning on the scale, we remedy this by converting each measure to a *z-score* using the estimated mean and standard deviation. Although

¹⁸The instructions provided to annotators and the annotations are available at <https://zenodo.org/record/7641732> (DOI: 10.5281/zenodo.7641732). We note that the same protocol was also used to curate the automated topics found by MALLET in the previous chapter (Chapter 3).

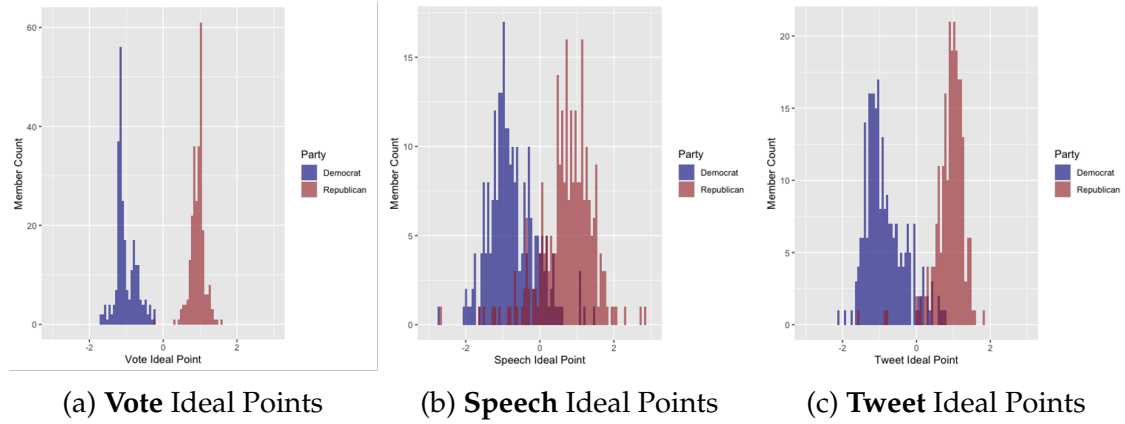


Figure 4.2: Member distribution of the three ideal point measures, 115th - 116th Congress.

this obscures differences in the relative dispersion of the measures — for example, it may be the case that text-based ideal points show greater variation than votes, perhaps indicating increased polarization of speech — in practice the distributions are similar. The z-score has the intuition of quantifying each legislator by how extreme they are relative to their peers per measure. The z-score-based process for comparing vote and text ideal points is drawn from [Vafa et al. \(2020\)](#).¹⁹

4.6 Results and Analyses

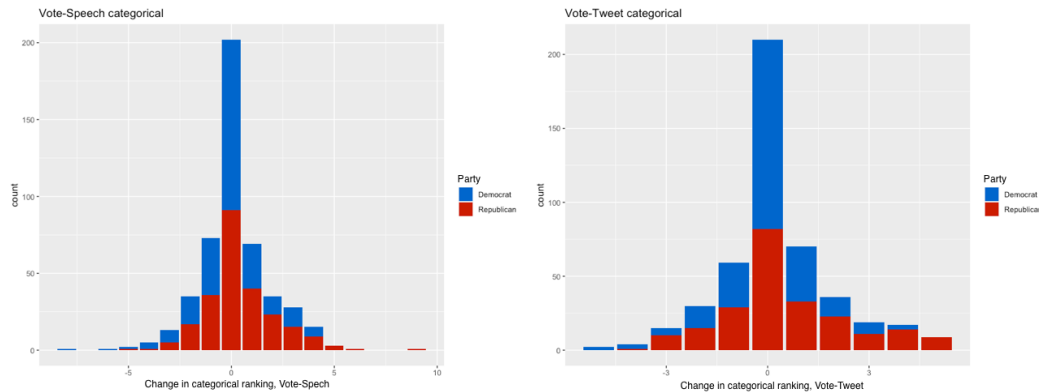
The Differences in the Three Ideal Point Measures. The result of this data collection and modeling are three ideal points — one for votes, one for floor speeches, and one for tweets from members across the 115th and 116th Congresses. Consistent with convention, negative scores are more “liberal” ideal points, and positive, higher scores are more “conservative”. The three ideal point measures capture different dynamics across the parties.

¹⁹While this ultimately makes vote and speech ideal points more easily comparable, because all information is not shared between each ideal point type (not all topics that are tweeted or spoken about receive a vote), the results should still be considered independently. We leave the estimation of a model that jointly estimates speech and votes to future work.

Fig. 4.2a shows the distribution of member ideal points using roll-call votes. As expected, the two parties are highly polarized, and within the two parties, vote-based ideal points are the most condensed of the measures. The median for the Democratic Party is -1.109 with a range of 1.483 , and the median ideal point among Republican House members is 0.962 , with a range of 1.856 . Given the centralized control over what proposals come to a floor vote, as well as the binary choice they present, there are few opportunities for individual members to express nuanced ideological positions, and hence Fig. 4.2a illustrates the concentration of ideal points by party.

The distribution of speech-based ideal points is showcased in Fig. 4.2b, and reveals that although the two parties are polarized, there is a notable overlap of members with more “moderate” ideal points. The median speech-based ideal point for both parties is also less extreme than when measured by votes: the median score for Democratic members is -0.867 , and for GOP members is 0.794 . Not only are the two parties closer together ideologically, but there is also greater variation in speech ideal points within the two parties: Democratic members are spread across a 4.143 range, and GOP members across a 5.572 range, indicating greater variation in how co-partisans position themselves ideologically in their speech as compared to votes. In other words, measuring ideal points through speech uncovers more conservative positioning by Democrats and more liberal positioning by Republicans than what’s identified by vote-based ideal points.

Lastly, Fig. 4.2c presents the distribution of tweet-based ideal points. Here we see two polarized parties, but with more within-party variation than was captured with vote-based ideal points, and less variation than was revealed with the speech-based ideal points. Reflective of this dynamic, the partisan medians are -0.994 for Democrats and 0.984 for Republicans, and the ranges are 2.912 and 3.371 re-



(a) Categorical shift from vote-based ideal points to floor-speech-based ideal points. (b) Categorical shift from vote-based ideal points to tweet-based ideal points.

Figure 4.3: Categorical shift in member rankings from vote-based to text-based (floor speeches and tweets) ideal points, 115th and 116th Congress.

spectively. This is consistent with Twitter as having both the potential to be less polarized because it is outside the partisan structures of Congress, but also the tendency to enable more extreme appeals to a national, polarized audience.

Altogether, our vote-based ideal points look like what we expect of Congress — polarized parties with tight intra-party cohesion and little overlap between the parties. Floor speech-based ideal points offer more space for ideological expression and show a Congress that is polarized, but less extreme with wide distributions of preferences within parties and space for moderates in both parties. Twitter-based ideal points present a Congress that also allows for more ideological expression, but that shows more partisan polarization than floor speech.

Explaining the Differences: Congressional and District Pressures on Parties (or the Whole Chamber). Given the differences revealed across the three ideal point measures, we examine how they reflect congressional and district pressures across the chamber (or party) as a whole. More specifically, we predict the impact of district, chamber, and individual considerations on each ideal point measure using

<i>Independent Variable</i>	<i>Dependent Variable</i>		
	Vote Ideal Point	Speech Ideal Point	Tweet Ideal Point
Female MC	0.0184 (0.025)	-0.168*** (0.065)	-0.084* (0.045)
Non-White MC	0.009 (0.028)	0.005 (0.069)	-0.097* (0.054)
Number of terms served	-0.014*** (0.003)	-0.002 (0.008)	0.005 (0.008)
Percent vote in last election	-0.001 (0.001)	-0.005 (0.005)	0.001 (.003)
District vote for Trump	0.007*** (0.002)	0.010** (0.005)	0.014*** (0.004)
Suburban district	0.037 (0.039)	-0.335*** (0.091)	-0.126 (0.078)
Urban district	0.052 (0.045)	-0.148 (0.113)	-0.196** (0.090)
District percent white	-0.001 (0.001)	-0.001 (0.002)	-0.0001 (0.002)
District unemployment rate	-0.053*** (0.016)	0.033 (0.031)	-0.001 (0.030)
Problem Solvers Caucus	-0.048 (0.032)	0.099 (0.092)	0.069 (0.068)
Progressive Caucus	-0.017 (0.028)	-0.226*** (0.074)	-0.184** (0.053)
New Democratic Coalition	0.039 (0.030)	0.054 (0.081)	-0.017 (0.063)
Blue Dog Coalition	0.333*** (0.046)	0.579*** (0.108)	0.504*** (0.090)
Party leader	-0.067* (0.036)	-0.233** (0.110)	-0.186** (0.085)
Committee chair	0.034 (0.050)	0.247* (0.134)	-0.030 (0.121)
Top committee member	-0.048** (0.022)	-0.200*** (0.064)	-0.101** (0.048)
Constant	-0.915*** (0.159)	-0.543 (0.442)	-1.140*** (0.324)
Observations	249	241	244
R ²	0.599	0.477	0.537

Table 4.1: **OLS regression results for Democratic party ideal points across venues, 115th – 116th Congress.**

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$; MC = Member of Congress.

ordinary least squares (OLS) regression models. Each ideal point measure across different venues is estimated separately and the analyses are separated by party for

<i>Independent Variable</i>	<i>Dependent Variable</i>		
	Vote Ideal Point	Speech Ideal Point	Tweet Ideal Point
Female MC	-0.023 (0.027)	-0.034 (0.075)	-0.115* (0.062)
Non-White MC	-0.020 (0.037)	0.067 (0.121)	-0.337** (0.149)
Number of terms served	-0.008*** (0.003)	-0.040*** (0.008)	0.008 (0.006)
Percent vote in last election	0.004** (0.002)	-0.002 (0.004)	-0.004 (0.003)
District vote for Trump	0.005*** (0.002)	0.010** (0.004)	0.011*** (0.003)
Suburban district	0.024 (0.028)	-0.225*** (0.081)	-0.012 (0.040)
Urban district	-0.051 (0.051)	-0.493*** (0.137)	-0.848*** (0.323)
District percent white	-0.003*** (0.001)	-0.005 (0.003)	-0.0001 (0.002)
District unemployment rate	-0.042*** (0.010)	0.035 (0.031)	0.054*** (0.021)
Problem Solvers Caucus	-0.093* (0.050)	-0.053 (0.078)	0.059 (0.059)
Republican Study Committee	0.072*** (0.029)	0.027 (0.071)	0.067 (0.043)
Freedom Caucus	0.063** (0.027)	-0.244** (0.105)	-0.030 (0.054)
Party leader	0.013 (0.030)	-0.061 (0.105)	0.166*** (0.040)
Committee chair	-0.033 (0.032)	-0.326** (0.144)	0.005 (0.050)
Top committee member	-0.028* (0.017)	-0.152* (0.087)	0.120*** (0.041)
Constant	0.801*** (0.155)	1.101*** (0.386)	0.253 (0.211)
Observations	257	243	227
R ²	0.440	0.318	0.481

Table 4.2: OLS regression results for Republican party ideal points across venues, 115th – 116th Congress.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$; MC = Member of Congress.

ease of interpretation.²⁰ At its most simple interpretation, a negative value means that the resulting ideal point is more liberal, holding all else constant. A positive

²⁰Full models that consider party as an interaction variable can be found in Appendix C.4, Table C.6. The results are consistent across specifications and we present the split sample here for ease of interpretation.

value reflects a more conservative ideal point. Thus, for the Democratic party, while a negative value indicates that members are reflecting a more liberal ideal than other members of the party, a positive value indicates a moderating effect on a member's ideal point – *i.e.*, more towards the “center” of the chamber. Conversely, a negative value for Republicans indicates a moderating influence towards the “center” of the chamber, while a positive value indicates a more conservative impact.

Tables 4.1 and 4.2 present the results of the Democratic party and Republican party models respectively. Across both parties, legislators' ideal points are largely determined by district and institutional variables and this is true whether measured by vote, speech, or tweets. Perhaps the most consistent finding is that increased district support for Donald Trump in the 2016 presidential election predicts more conservative ideological positions for both Democrats and Republicans.²¹ While there is some variation in exactly which district factors and institutional position predict ideal point measures, a striking difference is that legislators' gender and race are not predictors of *vote*-based ideal points, but do play a significant role in understanding ideological positions based on speech and tweets.²²

For Democrats, gender is an important determinant of ideological positioning as women are more liberal than their male colleagues all else equal. This suggests that female Democratic legislators use floor speech and tweets to establish positions that voting alone does not accommodate. This finding is consistent with previous work on female legislators' use of speech in the US and cross-nationally (*e.g.*, Vallejo Vera and Gómez Vidal, 2022; Pearson and Dancey, 2011b; Russell, 2021). Table 4.1 also reveals that non-white Democratic legislators take more liberal positions than their

²¹This finding for Republicans mirrors the results in Ballard et al. (2023), which show that members from ‘safer’ districts are more prone to sending polarizing tweets.

²²We use a binary definition of gender and an indicator variable to denote female legislators. For legislators' racial identity, we code all legislators as white or non-white, where non-white includes African American, Latino, Asian American and Pacific Islander, and Native American legislators.

white colleagues when looking at tweet-based ideal points, which indicates that minority legislators employ Twitter to carve out a more personalized ideological position whereas white legislators use the flexibility of Twitter to take a more conservative position (than their votes alone would indicate). These relationships also hold true among Republican legislators (Table 4.2) as female and non-white Republicans have more liberal tweet-based ideal points than their co-partisans, indicating the use of social media to differentiate oneself within the party. We note that the small number of legislators in these categories warrants caution in drawing conclusions about dynamics within the Republican party.²³

More generally, the examination of tweet-based ideal points reveals the factors that push legislators towards more ideologically extreme positions on Twitter. When looking at vote and floor speech-based ideal points, we find that some factors push legislators in a more extreme direction while other factors encourage moderation. But when examining tweet-based ideal points, the factors that affect legislators' positions overwhelmingly push towards greater ideological extremism (e.g., Republicans being more conservative and Democrats being more liberal). There are just two notable exceptions to this pattern: Republicans from more urban districts take more moderate positions than their colleagues and Democrats from districts highly supportive of Donald Trump's 2016 presidential campaign also take more moderate positions. The pattern of more ideologically extreme positions is consistent with Twitter's more national audience, as well as the general tendency of social media to encourage extremism and partisanship (Hong and Kim, 2016; Sunstein, 2018; Yarchi et al., 2021).²⁴

²³There are 24 female and 10 non-white Republican legislators in the dataset.

²⁴We note that multiple studies find and encourage a more nuanced view of social media's role in polarization; see Barberá (2020) for an overview. O'Hara and Stevens (2015) found counter-evidence to the narrative of the Internet (broader than but including social media) contributes significantly to violent extremism.

	Democrats		Republicans	
	Shift Vote to Speech	Shift Vote to Tweet	Shift Vote to Speech	Shift Vote to Tweet
Female MC	18.709 (11.683)	15.573* (8.846)	-12.246 (15.854)	10.324 (20.268)
Non-White MC	6.379 (14.557)	17.029* (10.425)	-12.593 (22.322)	59.155* (33.253)
Number of terms served	-3.062** (1.530)	-3.739*** (1.365)	3.547*** (1.271)	-6.424*** (1.485)
Vote in last election	0.337 (0.535)	-0.433 (0.502)	1.273** (0.658)	1.408 (0.927)
District vote for Trump	0.452 (0.971)	-0.396 (0.761)	0.311 (0.930)	0.638 (1.099)
Suburban district	50.364*** (16.591)	25.325* (14.531)	31.920** (13.281)	6.871 (13.140)
Urban district	32.234 (21.347)	40.091** (17.169)	49.505** (23.793)	81.856 (59.217)
District percent white	0.336 (0.455)	0.095 (0.395)	-1.031* (0.553)	-2.131*** (0.581)
District unemployment rate	-12.127 (7.516)	-6.720 (6.771)	-27.265*** (4.980)	-36.407*** (5.916)
Problem Solvers Caucus	-28.163** (12.594)	-25.213*** (10.183)	-15.670 (16.037)	-32.341** (16.001)
Progressive Caucus	21.389* (12.594)	14.242 (11.231)		
New Democratic Coalition	7.381 (14.176)	13.615 (11.509)		
Blue Dog Caucus	4.293 (18.269)	7.683 (9.532)		
Republic Study Committee			24.492** (11.875)	9.782 (13.963)
Freedom Caucus			73.036*** (15.825)	29.549* (16.724)
Party leader	11.264 (18.269)	10.630 (17.525)	25.750 (19.315)	-27.521* (14.775)
Committee chair	-40.698* (24.282)	0.741 (23.602)	14.792 (21.978)	-25.665 (16.521)
Top committee member	10.535 (10.590)	5.448 (8.625)	2.787 (12.265)	-44.748*** (11.841)
<i>Constant</i>	-58.703 (73.826)	30.130 (62.401)	33.481 (68.359)	221.295 (68.070)
Observations	246	244	243	227
R ²	0.188	0.207	0.322	.408

Table 4.3: OLS regression results for change in ranked ideal point positions across venues, 115th – 116th Congress.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$; MC = Member of Congress.

The Impact of Individual Legislator-level Factors. Given that chamber-wide trends across the three measures of legislative ideal points indicate ideological differentiation across our three mediums, we now turn to the question of which *individual* legislators shift their ideological positioning based on the venue, and under what conditions. We expect legislators to deliberately deviate from their vote-based positions when necessary to more accurately reflect their ideal point. Floor speeches and tweets both allow legislators to adjust their positions in order to differentiate themselves from their party and to take a more personalized position. The motivations for such fine-tuning of ideological positions may be rooted in constituency representation and the electoral incentives for legislators to more accurately reflect their district. Legislators may also use speeches or tweets to more fully reflect their own preferences, or their position (or aspirations) within the institution. Our analyses allow for all of these motives to play a role.

In order to evaluate individual-level shifts in ideological positions across venues, we first address comparability across the three ideal point measures. Since the three ideal point estimates are generated using independent models and are not directly comparable, we remedy this by comparing legislators' relative positioning on the scale for each measure. For each ideal point measure, we rank legislators by ideal point. This produces an ordinal ranking for members in order of liberal to conservative positions based on vote-based ideal points, another ranking based on speech-based ideal points, and a final ranking based on tweet-based ideal points. We use these rankings to evaluate which members are more or less conservative (or liberal) in how they talk or tweet, compared to how they vote. For example, Rep. Kevin Brady (R-TX), the Republican chair of the Ways and Means Committee, is ranked as one of the more conservative voters in the chamber (413), but in his floor speeches to his peers, he is 70 points more moderate (343). Yet when he addresses his

national audience on Twitter, he presents a more conservative ideological leaning (ranked at 470).

In order to more easily depict the degree of relative movement in ideal point positions, we generate a categorical variable that captures how far an individual member shifted in their ideological presentation. Members who shifted more than 0.5 standard deviations (44 spots), are coded as 1 (or a one-unit shift in a liberal/conservative direction). Members who shift one standard deviation (88 spots) are coded as a 2, or a two-unit shift in a liberal/conservative direction, and so on. We use this categorical variable in Fig. 4.3 to illustrate how prevalent changes in relative ideological positions are at the individual legislator level.

In categorical shifts shown in Fig. 4.3, a positive shift is indicative of a liberal direction (a more conservative vote-based ideal point minus a more liberal text-based ideal point), while a negative shift is indicative of a more conservative shift. Each categorical increase (decrease) is a half-point standard deviation increase (decrease), towards a liberal (conservative) direction. As Fig. 4.3 indicates, members are most likely to stay put ideologically — their text-based ideal points, derived from both floor speeches and tweets, are similar to their vote-based ideal points.²⁵

However, even with this conservative metric, there is still a great deal of variation in ideological presentation among members. The vote-to-speech ideological shift (Fig. 4.3a) is more widespread and common than the vote-to-tweet ideological shift (Fig. 4.3b). The two parties use text-based ideal points to moderate their vote-based ideal points, with Republican members more likely to move in a liberal direction and Democratic members in a conservative one. However, Republican members show greater variation in their ideological presentation than their Democratic

²⁵We recognize that even a “zero” ranking likely captures a slight shift (< 44), but we consider this relatively stable, particularly given the normal distribution of our categories.

counterparts. Republicans are more likely to use speeches and tweets to present a different ideological position — both in a liberal and conservative direction.²⁶

To examine why legislators shift their ideological position across votes, floor speeches, and tweets in greater detail, we estimate the change in legislators' relative ideological rank within their party. We use this to compare vote-based ideal point rankings to each alternate measure: speech-based ideal points and tweet-based ideal points. For each model, we evaluate the impact of individual-level factors such as legislators' identity and seniority in the chamber, district considerations such as electoral security, partisanship, and composition of the district, and institutional considerations such as leadership positions, membership on top committees, and party caucus affiliations. This helps determine the conditions under which legislators are more liberal or more conservative when measured by speech or tweets as compared to vote behavior. The results of these analyses are presented in Table 4.3 and reveal how legislators might be using behaviors other than voting to exhibit different ideological positions relative to their party members in order to reflect their district and institutional roles.

The analyses in Table 4.3 reveal several important dynamics. First, the type of district a legislator represents is important in determining how that legislator positions themselves through floor speeches and tweets. Notably, we find consistent evidence that urban and suburban districts drive legislators to express relatively more liberal positions than those captured by vote-based ideal points. This is consistent with cities generally leaning more liberal and Democratic, but the dynamics of the suburbs are very much a contested political landscape and both recent academic and journalistic accounts suggest movement in the liberal direction (Samuels, 2022; Tam Cho et al., 2013). This liberal shift means that Democrats in urban and

²⁶See Appendix C.4 for individual-level shifts, sans categorical deviations (Figs. C.1a and C.1b).

suburban districts shift towards the ideological extreme of their party, but when Republicans represent these districts, the same shift means a more moderate position or movement towards the ideological center. It is also worth noting that the moderating effect on Republicans only exists when looking at speech-based ideal points, and the urban or suburban nature of the district does not compel a more liberal (moderate) position in their tweets.

For Republicans, the racial composition and economic conditions of their district also shape how they position themselves when comparing votes to speech and tweets. Republican members who represent districts with a higher percentage of white residents, as well as those from districts with higher unemployment rates lead legislators to talk in ways that communicate a more conservative ideological position relative to their colleagues than their voting behavior would indicate. This shift is consistent with the rise of populist messaging on the political right and the use of racist undertones in some conservative politics (Kelly, 2010; Wilson, 2020; Hartman et al., 2021). While normatively problematic, this shift is expected given the different — and increasingly partisan — audiences across the three venues. The effect of unemployment rates on the likelihood of conservative messaging also echoes a well-established relationship between Republicans and economic anxiety, genuine or not (Pierson, 2017; Hacker and Pierson, 2005)). It is also consistent with the institutional constraints on roll-call voting (*e.g.*, leaders' agenda-setting powers, the Hastert rule,²⁷ and more) that promote less extreme ideological expression through voting and therefore contribute to the need for legislators to stake out an ideological position that votes do not capture by shifting how they position themselves (or shifting the perception of their positioning) through their language.

²⁷"The Hastert Rule says that the Speaker will not schedule a floor vote on any bill that does not have majority support within their party—even if the majority of the members of the House would vote to pass it." Source: https://en.wikipedia.org/wiki/Hastert_Rule.

Legislators' own lived experience and their experience in the chamber affect how they use language to differentiate their position from their votes. More senior legislators of both parties use floor speeches to position themselves more moderately than their voting behavior, such that Republicans move in a more liberal direction and Democrats in a more conservative direction. This moderating effect of seniority is repeated among Democrats when comparing their vote-based ideal points with their tweet-based ideal points. However, more senior Republicans actually take *more* conservative positions through their tweets than their votes. The finding that more senior legislators moderate their positions in their floor speech likely reflects the long-standing (but arguably declining) norms of the chamber, including civility, institutional loyalty, and coalition-building ([Matthews, 1959](#); [Hanges et al., 2020](#)).

Additionally, as suggested by the aggregate patterns, legislators' gender and race influence their positioning on Twitter as compared to their voting behavior. Among Democrats, women and non-white legislators shift their ideological position in the more liberal direction when comparing tweet-based ideal points to vote-based ideal points, consistent with work on legislative ideology and gender ([Reingold, 2008](#); [Osborn et al., 2019](#); [Thomsen and Sanders, 2020](#)). Similarly, non-white Republican legislators also shift their ideological position in a more liberal direction when comparing Twitter to roll-call votes.²⁸ These findings suggest that, at least for Democratic women and minority members, party leadership might not be bringing to vote the bills and propositions that sufficiently allow the expression of their ideological positions, and so they use other forums outside the constraints of Congress (such as Twitter) to convey these positions.

Finally, our data show that institutional position drives some shifts in ideolog-

²⁸Again, a caveat here is warranted given the small number (10) of non-white Republicans who served in the House during this period.

ical positions. Notably, intra-party caucuses are important predictors of shifts in legislators' relative ideological positions, and this is more pronounced for Republicans than Democrats (Gaynor, 2021). Republicans who belong to the Republican Study Committee (RSC) or the Freedom Caucus, two conservative caucuses in the chamber, moderate their positions in their speeches and take relatively more liberal positions than their votes convey. This finding is somewhat unexpected given the reputation of conservative House members for engaging in ideologically extreme rhetoric. While further examination is called for, one consideration is that conservative Republican legislators stake their identity on their largely oppositional *votes* to the Democrats' legislative agenda, even on bills that receive bipartisan support. In this way, their vote-based ideal points may sufficiently capture their conservative ideological position, allowing them to moderate slightly when speaking on the floor.

For Democrats, there is evidence that members of the Progressive Caucus shift to take a more liberal position in the floor speech than their voting records capture, but this dynamic does not carry over to their ideological expressions on Twitter. Nor do we see evidence of membership in other intra-Democratic-party caucuses (such as New Democrats or Blue Dog Caucus) affecting whether legislators shift their ideological positions when measured through speech or tweets. In contrast, Democratic members of the explicitly bipartisan Problem Solvers Caucus take steps to present a more moderate ideological position in both floor speeches and tweets compared with votes. This suggests that although they may vote with their party, they use other venues to differentiate themselves and establish a more moderate position consistent with the Problem Solvers Caucus. However, membership in the Problem Solvers Caucus does not have this effect for Republicans who position themselves as more conservative on Twitter than their vote-based ideal points.

While inconsistent with the goals of the Problem Solvers Caucus, the national audience of Twitter may create incentives for Republican members to establish more conservative positions for an audience outside of Congress even while espousing principles of bipartisanship within the chamber.

Taken together, these analyses illustrate how individual legislators take different ideological positions across votes, floor speeches, and tweets. This deliberate adjustment to their relative ideological position throughout the chamber reflects legislators' incentives to adapt to these distinct venues that have their own rules, audiences, and legislative implications. The factors that explain why some legislators are more or less liberal in floor speeches or tweets as compared to their votes highlight the importance of constituency considerations and the importance of their institutional position. As expected, the more restricted nature of voting — both in the control over what comes to a vote as well as the limited choice of voting for or against — means that sometimes it cannot capture all the nuance of legislators' ideological expression. Legislators, therefore, use other opportunities such as floor speeches or communicating on Twitter to express their ideological positions, often deviating from their vote-based positions.

4.7 Discussion and Conclusion

To many observers of Congress, the conclusion that legislators behave differently across different settings is not surprising nor controversial. Indeed, what these data reveal is conventional wisdom on Capitol Hill. However, the academic study of legislative ideal points has sometimes assumed a more fixed view of ideology where vote-based ideal points capture the singular underlying position of a legislator. We contend that they are one of the multiple possible measures of ideological position-

ing, and no single latent variable can capture a ‘true’ or fixed position in a valid manner. Looking at a legislator’s ideal points across venues and understanding the shifts in positioning provides a more complete picture of how legislators express their preferences to colleagues and constituents.

The addition of text-based ideal points also expands scholars’ ability to evaluate ideological representation, as well as examine ideological positions of those who do not have voting records such as first-term members of Congress.

From a methodological perspective, the ideal points developed here rely on more data points (over 435 House members versus 100 Senators) and over a longer time span (two congressional sessions versus one) than prior work by [Vafa et al. \(2020\)](#). Although this results in less-structured models than prior iterations, we find cohesive and replicable topics across speakers. Notably, even for short, social-media posts, we derive reliable estimates of ideology. In addition to measuring modeling stability computationally, we also assess the reliability of the text-based ideal point method via original, extensive human annotation and validation efforts. Ultimately, our models automatically infer high-quality ideal point estimates by authors. The automated discovery of topics and frames across ideology is a notable and open challenge in computational text analysis and computer science research more generally. Our research shows how improvements in the computational domain can shed new light on theoretical questions. Of course, like all work that uses text as data, questions about the authenticity of an ideological expression come to mind. But regardless of this caveat, we do believe that these new ideal point models present significant potential for the study of ideology in Congress (and beyond). The unsupervised setup allows for easy adoption on different datasets and contexts since only a collection of texts and their authors is needed.

Ultimately, these measures indicate that members of Congress shift their ideolog-

ical position depending on the venue used to express their ideology, and these shifts are reflective of both district pressures and institutional responsibilities. While legislators do not all move in the same direction, nor do they shift to the same degree, there are underlying dynamics to these individual-level adjustments in ideological positions. First, the different rules and constraints of each venue shape the expression of ideological position. Second, we find evidence that legislators use these three venues to reach different audiences and express different ideological positions better tailored for their constituencies.

When looking at the individual-level data, captured by how much a legislator's ideological position across venues changes when compared to their peers, we find evidence that lawmakers are catering to their district, and in some cases, the institution. At the district level, we find that lawmakers representing urban and suburban areas express more liberal positions in speeches and tweets. Thus, for Republicans representing these areas, speeches and tweets have a moderating effect on their voting records. Also, for Republicans, the demographic and economic makeup of their district impacts their ideological expressions: Republican members that represent majority white areas and areas of high relative unemployment are more conservative in how they speak to their constituents — a possible reflection of the populist direction of the Republican Party.

We also find evidence that institutional expectations impact ideological positioning. Senior lawmakers moderate their (likely party-line) votes when speaking to their peers, but still use Twitter to express more conservative positions. This is similar to Republican members of congressional caucuses, including the bipartisan Problem Solvers Caucus — while floor speeches were more moderate than votes, Twitter still provides Republican lawmakers a chance to be more conservative than their voting record allowed. This indicates the staying power of norms within the

chamber, juxtaposed against the increasing importance for Republicans to cater to the national base.

Lastly, we found that Democratic women and non-white members, and non-white Republican members expressed more liberal positions on Twitter than their roll-call votes. Like the other findings reiterated here, this is indicative of the limitations of roll-call votes to express an ideological position. Personal attributes, district pressures, and institutional positions lead lawmakers to use floor speeches and social media to differentiate themselves ideologically. When only given the choice of ‘yes’ or ‘no’, text-based expressions allow for greater nuance that lawmakers gladly take.

Notably, Republican members, across our models, were more likely to use these outlets to express conservative positions. Future work should build on this, considering how these findings interact with existing work on asymmetrical polarization between the two parties. As social media in particular becomes more impactful and important in the political dialogue of elites and constituents alike, understanding how lawmakers are using these outlets to encourage ideological extremism could potentially become more important than votes.

Scholars have long accepted that Congress is becoming more polarized, but the roll-call mechanism by which these assumptions are made is inherently limited. We have found that by using textual data, we can more insightfully capture the ideological positioning of a member of Congress. When comparing the ideal points of votes, speeches, and tweets, this produces notable differentiation that speaks to the influence of district and institutional pressures, as well as demographic attributes. These models and the ideal-point outputs they produce can easily be applied to other questions of ideology, even by policy area (we concretely present this idea as part of future work in Section 6.2).

With our extension allowing for topic outputs of models like MALLET to initialize TBIP, and our efforts establishing the model’s stability and reliability, an original dataset combining ideal point measures with various kinds of legislator information and the investigation we conduct in this study, we have demonstrated how an existing unsupervised NLP can effectively help validate existing hypotheses as well as contribute new knowledge in social and political science research.

4.8 Potential Limitations and Caveats

While the TBIP model tries to control for the length of documents by taking the *log* of counts instead of raw counts, another possible concern is the influence of the number of documents authored by someone (or their frequency of speaking on the floor or tweeting) on their ideal point estimates. It is important to disentangle if the model is learning a strictly political spectrum or if the frequency of speaking or tweeting is impacting the extremity of their ideal point values.²⁹ We conduct a preliminary analysis that suggests that this is not the case; the correlation between the number of documents authored by someone and the absolute distance between their ideal point value from the median ideal point is low (Spearman $r = 0.144$ for floor speeches and Spearman $r = 0.388$ for tweets). This result also suggests that tweets might be more prone to seeing an effect of the frequency of tweeting on learned ideal point values. Future work should look more into the possible effect of the number of texts authored on ideal point values, which may not be as simple as

²⁹A related concern that also applies to vote-based spatial models measuring latent traits is when people on the two ‘extremes’ show identical behavior but for considerably different reasons as thoroughly examined in [Duck-Mayr and Montgomery \(2022\)](#); the authors introduce a method to account for this different-motivations-for-extremity caveat for latent measurement and future work should attempt to extend their approach to text-based measures. We acknowledge that this effect of ‘extreme’ members is a limitation of our estimates in this work, and lack of distinct accountability of extremity effects could help explain some of the results such as Republicans in the Freedom caucus being more moderate in their floor speeches (Table 4.2).

ideal points for those who author more documents being more extreme.

In addition to the frequency of tweeting, a more general consideration for Twitter is the variation in how much members might care about social media-based messaging even if they have official accounts, and how that might impact the estimates. For example, in the context of the 2013 German Federal Election, [Quinlan et al. \(2018\)](#) note the clear heterogeneity in social media adoption of different politicians. In the US, a 2020 report by Pew Research finds near-universal adoption and high levels of social media activity for elected representatives but also notes that there is a small subset of elected members who account for the majority of social media activity and clearly dominate the various social media engagement metrics ([Van Kessel et al., 2020](#)).

One signal towards interest and active engagement might be the number of followers a member has on Twitter; future work should consider including social-media-specific independent variables for individual legislators such as follower counts in the kind of regression analyses we conduct in this chapter.

A limitation identified by the [Vafa et al. \(2020\)](#) in their original introduction of the method is also worth highlighting here since we are applying the model to explain the real-world effects of legislative and political behavior. Since TBIP operates on bag-of-words, the context around a word may not be captured appropriately including what the sentiment of the author is. As demonstrated by [Vafa et al.](#), this can sometimes lead to cases where someone mentioning and using a term used more by authors of opposing ideology makes them more moderate when they are in fact very negative in their stance towards that term.

Chapter 5: Consequences of Communicative Choices: Mainstream News Articles Co-Shared with Fake News Buttress Misinformation Narratives

Most prior and current research examining misinformation spread on social media focuses on false news reports (‘fake’ news). This paradigm fails to consider the potential, interesting role of content published by mainstream media (‘real’ news) on account of their factual accuracy. However, factual mainstream news can still support or be used to support false and misleading narratives on social media. In this chapter, we operationalize narratives using an existing unsupervised NLP technique and examine the narratives (the *communicative choices*) present in misinformation content.

We find that certain articles from reliable outlets are shared by a disproportionate number of users who also shared fake news on Twitter. We consider these ‘real’ news articles to be *co-shared* with fake news. We show that co-shared articles contain existing misinformation narratives at a significantly higher rate than articles from the *same* reliable outlets that are *not* co-shared with fake news. This holds true even when articles are chosen following strict criteria of reliability for the outlets and after accounting for the alternative explanation of partisan curation of articles. For example, we observe that a recent article published by The Washington Post titled “Vaccinated people now make up a majority of covid deaths” was disproportionately

shared by Twitter users with a history of sharing anti-vaccine false news reports.

Our findings suggest a strategic repurposing of mainstream news by conveyors of misinformation as a way to enhance the reach and persuasiveness of misleading narratives. We also conduct a comprehensive case study to help highlight how such repurposing can happen on Twitter as a *consequence* of the inclusion of particular narratives in the framing of mainstream news.¹

5.1 Introduction

Much scholarly and media attention has been lavished on the presence of misinformation in the contemporary information ecosystem (Che et al., 2018; Ha et al., 2021; Weeks and Gil de Zúñiga, 2021); however, news published by legitimate sources adjacent to misinformation is potentially far more impactful. Indeed, studies attempting to measure the scale of the spread of, and exposure to, fake news as part of the information ecosystem have found it to represent a very tiny part of people’s media consumption (Allcott and Gentzkow, 2017; Grinberg et al., 2019; Guess et al., 2019; Allen et al., 2020; Altay et al., 2022b). For example, per Allen et al. (2020), “fake news comprises only 0.15% of Americans’ daily media diet.” Thus, news that is *not* deemed as fake has a much larger audience, and understanding its interaction with fake news and potential role in the world of online misinformation is critical (Watts et al., 2021). In this chapter, we show that one explanation for certain content from legitimate news sources being shared in the same social media circles as fake news is the occurrence of shared narratives.

Most research studying fake news relies on lists of particular outlets deemed as

¹This chapter is based on work accepted and presented at the International AAAI Conference on Web and Social Media (ICWSM) 2023: **Pranav Goel**, David Lazer, Philip Resnik. “Mainstream news articles co-shared with fake news buttress misinformation narratives”. ICWSM 2023.

‘fake news outlets’ ([Allcott and Gentzkow, 2017](#); [Budak, 2019](#); [Bozarth et al., 2020](#); [Robertson and Mourão, 2020](#)), which can unwittingly obscure the potential role of news from mainstream (“non-fake”) outlets in online misinformation. Operating at the level of news producers can also ignore heterogeneity in the quality of news articles published by any particular outlet, and how outlets deemed legitimate in terms of journalistic practices and norms might still publish individual, factual stories that are quite misleading (or easily used as such).²

In this work, we extract and analyze the articles published by mainstream political news outlets (popular outlets *not* marked as regularly publishing false information) that are most likely to be shared in the same Twitter circles as fake news in our dataset, i.e., *co-shared articles*. Articles from these mainstream outlets are scored on how likely they are to be co-shared with fake news. We hypothesize that legitimate news stories that are co-shared with fake news stories often contain misinformation narratives, which are entity-action relationships that repeatedly occur in fake news sources or are present in statements known to be false. This is in opposition to legitimate news stories published by the same reliable outlets that are not co-shared with fake news. In brief, our hypothesis statement is: mainstream news stories that are disproportionately co-shared with fake news consist of a significantly higher presence of misinformation narratives.

For example, a recent article published by The Washington Post (WAPo) with the headline “Vaccinated people now make up a majority of covid deaths” was disproportionately shared by Twitter users in our panel who have a history of

²As an example of one type of heterogeneity in content published by the same outlet, [Green et al. \(2021\)](#) find that news articles published by an outlet with a particular partisan leaning (on aggregate) are selectively curated and shared in a cross-partisan manner, since the articles themselves can span the political spectrum.

sharing anti-vaccine-related articles published by fake-news outlets (Fig. 5.1a).³ The headline and story content of the WAPO article does not contain any falsehoods to the best of our knowledge. According to our hypothesis, an explanation for this mainstream article being co-shared with anti-vaccine fake news is that the article (which includes the headline) contains narratives such as ‘vaccine cause death’ — a popular narrative and belief in misinformation content (Sharma et al., 2022). Indeed, we find that anti-vaccine article content published by fake news outlets in our dataset contains vaccine-related narratives such as ‘vaccine kill people’, ‘vaccine cause harm’, ‘vaccine use remains aborted child’, and ‘vaccine spread virus’. This example illustrates that the presence of misinformation narratives might enable the repurposing of an otherwise factually correct mainstream article by fake-news conveyors on Twitter.

We investigate our hypothesis by deploying a tool for automatically extracting narrative structures from English text that are represented using narrative strings or labels. This provides us with narratives present in fake news articles as well as claims fact-checked to be false (Fig. 5.1b). Our experiments across a variety of misinformation narrative collections and types of reliable outlets validate our hypothesis (the testing procedure is visualized in Fig. 5.1c). The difference in the presence of misinformation narrative strings between co-shared and control articles (from the same set of mainstream news outlets) remains significant when accounting for partisan differences in the Twitter audience of the two groups of articles (Appendix D.2).

³The particular article used for this example is: <https://www.washingtonpost.com/politics/2022/11/23/vaccinated-people-now-make-up-majority-covid-deaths/>. Disproportionate sharing was assessed using a method similar to the one explained in Section 5.3. We note that the headline shown in Fig. 5.1a was the one originally published (and shared when we assessed the uptake) and was later modified.

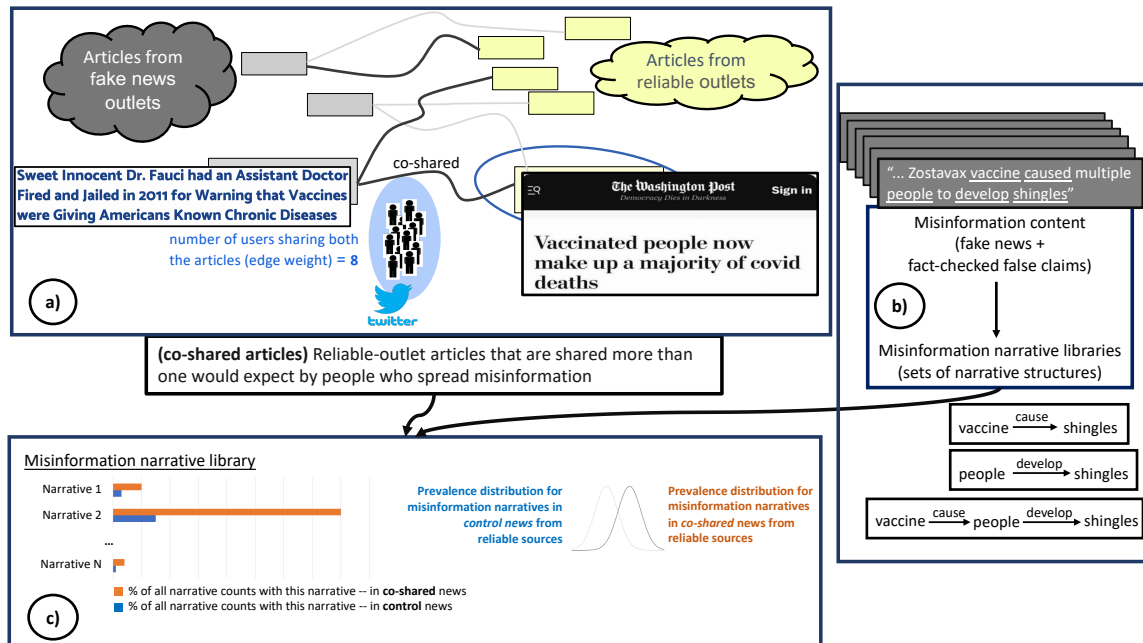


Figure 5.1: Visual illustration summarizing our approach. In part **a)** (top-left), we show our process of finding reliable articles disproportionately co-shared with fake news: we score pairs of fake and reliable news on the likelihood of co-occurrence of their shares (by the same set of Twitter users) when controlling for the individual popularity of the two articles. Our method uses a graph-based approach with edges between fake and reliable articles (the nodes) weighted by the number of users sharing both incident articles. The Washington Post article (only headline shown here) illustrates our approach: the article was disproportionately shared by anti-vaccine fake-news conveyors in our Twitter panel. This process helps us **create two groups of mainstream news articles: disproportionately co-shared and control**. Using a natural language processing method developed in prior work, we automatically extract potential narrative structures from all texts in our data: misinformation content as well as mainstream news articles, and **narratives from misinformation content are considered sets of misinformation narrative libraries** (part **b)**, right-hand side). Part **c)** (bottom-left) illustrates how we conduct the comparison at the heart of our **significance testing procedure, in order to test our hypothesis**: mainstream news stories that are disproportionately co-shared with fake news consist of a significantly higher presence of misinformation narratives.

Our method of scoring news stories on how disproportionately their Twitter shares are connected with fake news shares contributes a scalable and generalizable way of fleshing out the online misinformation ecosystem beyond fake news. It lays the foundations for further investigation of interacting mainstream and fake news articles.

Our finding has important implications for responsible journalistic practice: when vetting news stories (including the headlines) and their framing, it is important to consider not just the story itself but also the broader context of existing narratives in misinformation content. We further highlight the possible consequences of framing choices in mainstream news and their possible repurposing to support misinformation through a comprehensive case study, which also offers insight into how our work provides building blocks of an enabling technology for journalism (Section 5.5).

For both journalism and scientific research, our work highlights the importance of looking beyond falsehoods in fake news and towards the larger narratives and world views being fleshed out in news articles of all stripes, including legitimate news. These narratives, and not just the propensity to share falsehoods, are critical for understanding connections between fake news and other news articles.

5.2 Background and Definitions

While misinformation has existed since before the invention of the printing press (Burkhardt, 2017b), scholars have noted how online news consumption in the digital age, combined with social media, can enhance the speed and magnitude of its spread (Burkhardt, 2017a; Baptista and Gradim, 2020; Del Vicario et al., 2016; Vosoughi et al., 2018; Shu et al., 2017). Online misinformation spread has been noted for the multitude of crises it can help create, including the potential threat to the effective functioning of democracy (Persily, 2017; Aral and Eckles, 2019; Roozenbeek and Van Der Linden, 2019), the possibility of increased political polarization (Tucker et al., 2018; Kubin and von Sikorski, 2021), and negative effects on trust in public institutions as well as on interpersonal relationships (Lovari, 2020; Ognyanova et al.,

2020; Duffy et al., 2020).

However, as we noted above, fake news and online misinformation constitute a very tiny portion of people’s media diet and in this work, we aim to highlight the role of articles published by reliable (‘real’) outlets, with a much larger audience, in the world of online misinformation.

In this section, we define (and provide background for) the terminology used in this chapter. We use the terms *domains*, *sources*, and *outlets* interchangeably to refer to news outlets in this chapter.

Narratives. These are sets of entity-action relationships or structures that describe how entities act on each other (Sloman, 2005; Sloman and Lagnado, 2015). More specifically for us, these relationships are characterized by a base component of the following format: AGENT ENTITY $\xrightarrow{\text{VERB}}$ PATIENT ENTITY (Ash et al., 2021). Some examples are shown in Fig. 5.1b and Table 5.2; the extraction process is summarized in Fig. 5.3.

This is a specific instantiation using some of the components that have been found to characterize narratives in text across the broad and varied literature on narratives — Piper et al. (2021) review narrative theory and considered other components important to narratives such as timing, motivations, etc. Entities (agents and patients) and the actions connecting them were also found to be critical components of narratives in their review, and we focus on those following Ash et al. (2021), using their open-sourced tool for narrative extraction.⁴

Misinformation. This broadly refers to false information. We refer to content published by fake news outlets (per our particular operationalization described

⁴<https://github.com/relatio-nlp/relatio>

below) as well as claims or statements fact-checked to be false as misinformation in this work. The former is also sometimes referred to more specifically as fake news articles or fake news, and is typically a subset of misinformation.

Fake news outlets & fake news. Following [Lazer et al. \(2018\)](#) and [Grinberg et al. \(2019\)](#), fake news domains or outlets are the sources that masquerade as legitimately produced news but eschew epistemological practices (including editorial norms and processes for vetting information).⁵ In this chapter, we deem articles from outlets known to regularly publish false content (per annotations from NewsGuard ([NewsGuard, 2021](#))) as fake news, and the outlets themselves as fake news outlets.⁶

We note that prior literature has used NewsGuard annotations in different ways, such as using the provided ‘trustworthiness’ scores to divide information sources into groups ([Guess et al., 2021](#); [Edelson et al., 2021](#)). We aim for precision by considering ‘regularly publishing false content’ as our criterion out of all the factors that determine trustworthiness in NewsGuard ratings. This helps us highlight the drawbacks of conceptualizing problems with the digital information ecosystem as that of ‘false information’ or fake news content — a framework that often foregoes closer examination of information coming from sources deemed reliable. An example of a fake news article is shown in Fig. 5.1a (only the headline).⁷

⁵This definition is closely related to the usage of ‘the deliberate creation of pseudojournalistic disinformation’ in [Egelhofer and Lecheler \(2019\)](#).

⁶NewsGuard domain or outlet-level ratings and annotations have been used in prior work on understanding and quantifying misinformation ([Guess et al., 2021](#); [Edelson et al., 2021](#); [Altay et al., 2022b](#)). The independent organization consists of a team of journalists evaluating the quality of online information sources, and their domain-level judgments show a high overlap with other lists with publisher ratings such as the one used in ([Allcott et al., 2019](#)); specifically, recent work comparing six sets of news domains’ quality ratings (including NewsGuard ratings) found that “they generally correlated highly with one another” ([Lin et al., 2022](#)). Further information on NewsGuard ratings, comparisons with other evaluations, and validation can be found in [Guess et al. \(2021\)](#); [Edelson et al. \(2021\)](#). More details about the organization can be found at <https://www.newsguardtech.com/solutions/newsguard/>.

⁷Prior literature that did not use NewsGuard annotations also followed differing labeling schemes

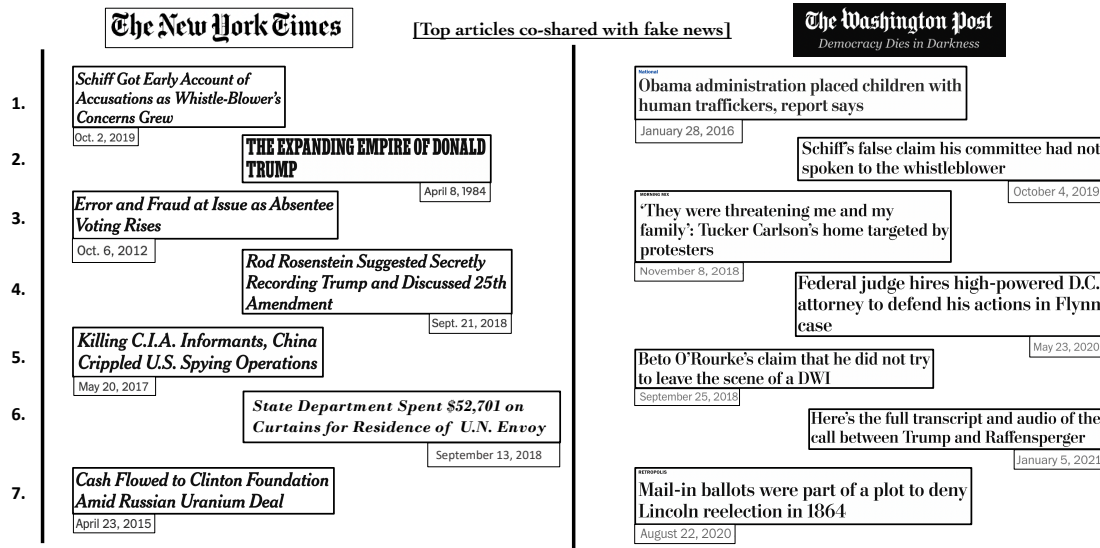


Figure 5.2: The top articles from The New York Times (NYT) and The Washington Post (Wapo) in the set of articles disproportionately co-shared with fake news, based on the co-sharing likelihood derived from Twitter sharing behavior. As motivation for our hypothesis in this work, consider the third NYT article and the seventh Wapo article shown in the figure about mail-in ballots and absentee voting. These articles are popular in the same Twitter circles sharing stories published by fake news outlets containing narratives (extracted in this work) that include ‘mail ballot raise risk fraud’, ‘lot people cheat mail’, ‘mail balloting increase incident fraud’, ‘paper mail balloting mean dramatic increase fraud’, and ‘republican complain potential mail absentee ballot fraud’. While only headlines are shown here for illustrative purposes, our investigation also utilized the entire text of the news article. We provide the texts as well as the exact article URLs for all groups of news articles as part of our codebase.

Reliable outlets. These are all the news outlets annotated for their quality as information sources by journalists at NewsGuard and *not* deemed as fake news outlets as per our selected criteria above. Since this is quite broad, we also show our primary results with narrower criteria such as only considering sources marked as ‘trustworthy’ by the NewsGuard team as well as ideological subsets (the process for liberal versus conservative categorization is described below).

for fake news sites. For example, Zimdars (2016) considered news sites that published completely fabricated stories as *fake* and also considered *unreliable* or *clickbait* as a separate category, while Bozarth and Budak (2021) combined all those outlets into one *low-credibility* category.

Mainstream media or news (outlets). Within reliable outlets, we define mainstream outlets as those that *publish predominantly political content*, are *popular*, and are *not ideological outliers*.

We use domain-level ratings as created by and used in [Green et al. \(2021\)](#). These ratings are based on Twitter shares of articles (over the 2010 to 2021 period) published by various outlets using the same panel of users in the overall dataset used in this work (briefly described in Section 5.3; the annotations we use are made available as part of our codebase).⁸ The three criteria we use to demarcate mainstream outlets are detailed below:

1. **Publish predominantly political content.** Each domain published many articles, but only some of them were deemed as ‘political’ in domain-level ratings based on the type of users sharing them on Twitter. Using the percentages of all articles published by a domain deemed political content as the political score, *predominantly political domains* are ones having a score greater than a z-score of 1.0 (greater than the mean plus the standard deviation calculated using all the domain-level political scores).
2. **Popular.** Among these domains that predominantly publish political content, we select the *popular* ones as those that are *i*) in the top 10% for the number of Twitter users sharing a domain’s content; and *ii*) among the top 500 most-visited domains per the users involved in a YouGov panel survey (2015-16), with site-visiting statistics derived from passive metering data (per the analysis and data-release by [Guess \(2021\)](#)). Using this intersection is a more robust way of deciding the domains that should be deemed ‘popular’ as opposed to, say, only considering popularity based on Twitter shares.

⁸https://github.com/Pranav-Goel/misinfo_narratives

3. **Not ideological outliers.** Finally, using the scalar aggregated ideological scores derived using the number of registered Democrat or Republican Twitter users sharing an article published by that domain (Green et al., 2021), we remove domains that are either in the top or the bottom 10% of ideological scores to remove ideological outliers.

The remaining outlets are referred to as *mainstream political news outlets*, *mainstream news*, or *mainstream media* in this chapter.

Trustworthy outlets. These outlets are a particular subset of the set of mainstream media outlets defined above — ones that are annotated as ‘trustworthy’ in NewsGuard ratings. We note again that ‘regularly publishing false content’ — our criterion for deciding the outlets that are considered as fake news outlets in this study — is one of several criteria that determine trustworthiness scores in NewsGuard ratings.

Liberal news outlets. Using the same aggregated ideological scores based on Twitter shares (Green et al., 2021) described above for removing domains that are ideological outliers, we consider each mainstream outlet with an ideological score less than the median of all scores across mainstream outlets as a liberal news outlet (the scores follow the traditional *negative-to-positive* as *liberal-to-conservative* axis, which means lower scores represent a more liberal rating).

Misinformation narratives. These are the entity-action relationships present in false claim statements or recurring (top 1% in terms of frequency of occurrence) in content published by fake news sources (such as “vaccines harm people”).⁹ The

⁹For the sake of being comprehensive, we include results where *all* structures extracted from fake news content are also used to test our hypothesis, not just recurring ones. But we define

method for extracting such narratives is detailed in Section 5.3 (and summarized in Fig. 5.3), and some examples are provided in Table 5.2.¹⁰ Misinformation narratives need not be exclusive to either misinformation or fake news, but in fact, as our findings indicate, can show up across both fake and ‘real’ news, cross-cutting the types of news outlets.

5.3 Data and Methods

Original Twitter Data. Using a dataset of English tweets from a nationally representative panel of U.S. Twitter users (Hughes et al., 2021; Shugars et al., 2021), we look at all tweets that share news articles (published by English-language information sources or outlets per NewsGuard) from May 2018 to November 2021.¹¹ This results in an initial dataset containing ~ 14.6 million unique URLs published by ~ 4000 domains shared by $\sim 600,000$ different users. For each tweet, we have the user ID of the author, and the news URL being shared in the tweet. We keep URLs that were shared by at least 20 different users (to control for sparsity effects), resulting in $\sim 420,000$ different URLs published by ~ 2400 domains shared by $\sim 480,000$ different users.

misinformation narratives that either occur repeatedly in fake news content (where each instance, in trying to mirror mainstream news, contains a detailed story spanning multiple sentences), as well as all the ones that occur in false claims (where each instance is a fact-checked false claim consisting of a short sentence).

¹⁰We note that narrative extraction, especially as captured in the narrative label or string representation, is noisy as can be expected from unsupervised methods. We discuss the limitation and potential future work in Section 5.7, and also provide all narratives extracted from false claims (that are used to test our hypothesis) in Appendix D.7 for the reader’s viewing.

¹¹The dataset used in this study was previously approved for research (for prior work; see (Hughes et al., 2021; Shugars et al., 2021)) by Northeastern University’s Institutional Review Board; we are not making use of any de-anonymizing data, and only use publicly available tweets and published news articles texts.

Creating two groups of mainstream news articles: disproportionately co-shared and control. Our process for finding co-shared mainstream articles is visualized in Fig. 5.1a.

We create a weighted graph where the nodes are individual articles or URLs shared in tweets (boxes in Fig. 5.1a), and edges are only considered to be between a pair of nodes where one node is an article published by a fake news outlet (nodes shown on the left-hand side of Fig. 5.1a) and the other node is an article published by one of the other (reliable) outlets (nodes shown on the right-hand side of Fig. 5.1a). Our processed dataset contains $\sim 24k$ fake news stories published by 276 domains. The weight of each edge is the number of Twitter users in our dataset that shared *both* the incident node URLs at least once (in our illustrative example in Fig. 5.1a, 8 is the edge weight since eight different users in our panel shared both the WAP0 and the fake news article displayed).

Each edge then receives a co-sharing score based on the likelihood of both incident node URLs getting shared by the same user, controlling for the individual popularity of each of those URLs in isolation. This is a straightforward application of a graph-pruning algorithm (Dianati, 2016) that has been applied in prior work to construct a co-exposure network that identifies news outlets with a shared audience ((Grinberg et al., 2019); also see Majó-Vázquez et al. (2019)). To the best of our knowledge, this method has not been previously applied to study networks formed by individual news articles. We posit that studying these networks of article-sharing is important for identifying mainstream news stories published by reliable outlets that would otherwise not be considered for their potential role in misinformation circulation on social media.¹²

¹²The complete implementation of the method is provided as part of our codebase at https://github.com/Pranav-Goel/misinfo_narratives.

Edges in the top 1% of co-share scores are considered disproportionately co-shared or simply, co-shared (highlighted as the darker edges in Fig. 5.1a), and any nodes (articles) from reliable outlets that are a part of the two incident nodes forming those edges are considered as co-shared.¹³ Some examples of top co-shared articles (based on an aggregated co-sharing likelihood score for each outward edge from a particular article or URL) from two particular news outlets are shown in Fig. 5.2.¹⁴ In this study, we only analyze articles published by mainstream outlets (as defined in the prior section) that fall in this top 1% set of articles as *the group of articles that are co-shared with fake news*.¹⁵ Stories published by the same set of mainstream outlets that are *not* in the top 5% of co-sharing likelihood scores form our *control group*. Finally, we have $\sim 91k$ news stories in the co-shared group and $\sim 51k$ news stories from the same set of news outlets in our control group.¹⁶ Headline and story texts of all URLs are automatically extracted.¹⁷

¹³Following recommendations and work in Dianati (2016); Grinberg et al. (2019), we set a “high threshold for deviation from the null model.” In order to get the 0.99 quantile (or top 1%) value to set the threshold, we draw 1000 samples of 1000×1000 sized subgraphs from the entire graph and use the distribution across the samples to estimate the threshold value with 95% confidence. The sampling procedure is used for estimation since using the entire $\sim 420,000 \times \sim 420,000$ graph was computationally intractable.

¹⁴We are only showing the top articles that are not opinion pieces, but we note that especially for WAPO, the top co-shared articles include opinion pieces.

¹⁵We note that mainstream outlets form more than 90% of all reliable outlets resulting in about 2000 domains publishing a total of $\sim 395k$ URLs in our dataset.

¹⁶We emphasize that the control group contains mainstream articles that do not have an outgoing edge to even a single fake news article with a co-sharing likelihood in the top 5% of all likelihood scores. This is a strict control criterion, separating the co-shared and control sets of articles, and explains the lower number of control articles (as opposed to what would be expected when simply considering every other article not in the co-shared set as the control group).

¹⁷We scrape contents of news articles using the URLs to get the headline, subheading, and story texts along with the publication date using the NEWSPAPER3K API, available as a Python library (<https://newspaper.readthedocs.io/en/latest/>). Our code as well as our entire collection of texts is provided as part of our codebase, though the Twitter data itself will be shared in a restricted manner following Twitter and IRB guidelines.

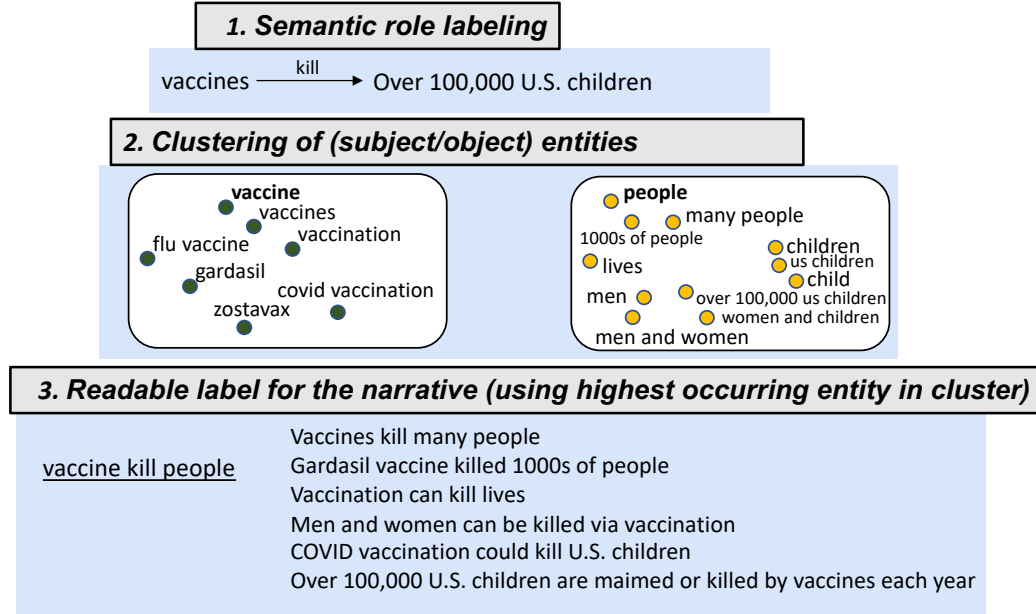


Figure 5.3: A brief summary of the process of extracting narratives using the method from Ash et al. (2021). The following example sentence, obtained from a text of a fake news article in our dataset, is used: “Over 100,000 U.S. children are maimed or killed by vaccines each year.”

Extracting and creating misinformation narrative libraries. As briefly explained in Section 5.2, we operationalize narratives using the methodology and open-source package created by Ash et al. (2021).¹⁸ The conceptual definition used is “sets of relationships between entities that act on each other”, which, in language, involves looking at “grammatical statements describing actors, actions, and the acted-upon.” Fig. 5.3 briefly visualizes the process of extracting the narrative structure through an illustrative example; while Fig. 5.4 provides the exact flowchart of the implementation of the open-sourced Python package (RELATIO) for narrative extraction and is taken directly from the manuscript created by Ash et al. (2021).

Narrative structures extracted consist of the base form:

$$\text{AGENT ENTITY} \xrightarrow{\text{VERB}} \text{PATIENT ENTITY} \in E \times V \times E$$

¹⁸<https://github.com/relatio-nlp/relatio>

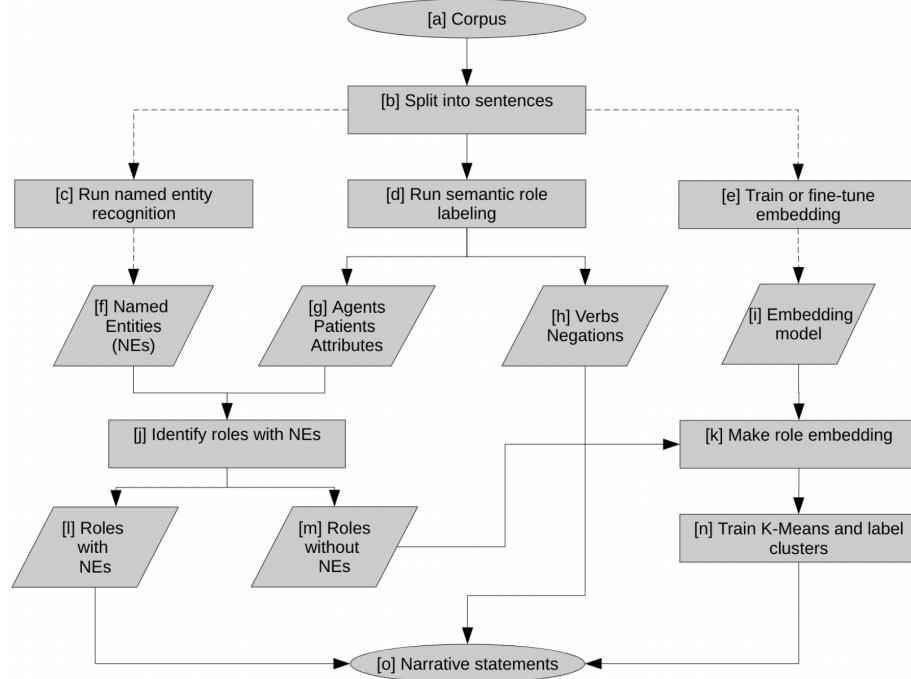


Figure 5.4: Complete flowchart of the exact implementation of narrative extraction in the open-source Python package: RELATIO (<https://github.com/relatio-nlp/relatio>); the image is copied directly from Ash et al. (2021).

where E is the set of clustered entities and named entities, and V is the set of all verbs. For example, $vaccine \xrightarrow{cause} shingles$ and $people \xrightarrow{develop} shingles$ could be extracted as these base structures, and we can also have $vaccine \xrightarrow{cause} people \xrightarrow{develop} shingles$ (extracted narrative structures need not be limited to triplets).¹⁹ These are then simply represented as strings, or what we call narrative labels or narrative strings (‘vaccine cause shingles’; ‘vaccine cause people develop shingles’).²⁰

¹⁹Example sentence used here is from the sentence “Zostavax vaccine caused multiple people to develop shingles.”, present in one article in our set of fake news. After the entire procedure such as clustering of multiple entities to be represented by one entity, the narrative $vaccine \xrightarrow{cause} shingles$ was extracted.

²⁰We note that the automated output of this package can be noisy, and not all extracted strings themselves will correspond to a meaningful notion of ‘who did what to whom’ — Ash et al. (2021) refer to the automatically extracted statements as narrative *candidates* for this reason. We will refer to these outputs as narrative strings or narrative labels, and sometimes simply as *narratives* in

Since entities are clustered together during the process of narrative extraction (Fig. 5.3), the pre-specified number of clusters constitutes an important decision. Following the documentation in the publicly available package by Ash et al. (2021), two settings are used in order to extract narratives at two levels of granularity from the text data. The first is called *low-dimensional narratives*: when fewer clusters or latent entities are allowed; and the second is called *high-dimensional narratives*: when a higher number of clusters or latent entities are allowed.²¹ Low-dimensional narratives are ‘broader’ and carry the risk of less coherent entities in these structures (since each group or cluster collects a larger number of underlying individual entity instances), whereas high-dimensional narratives are more ‘fine-grained’ and carry the risk of some clusters being redundant (and representing the same entity). For example, for the (original, non-processed) sentence — “The filing claims that Zostavax caused multiple people to develop shingles.” — the method finds the low-dimensional narrative ‘vaccine cause shingle’, and finds the high-dimensional narrative ‘zostavax cause multiple people develop shingle’. In the former case, the terms ‘zostavax’ and ‘vaccine’ terms were clustered together as part of the same entity cluster (represented by ‘vaccine’), while in the latter case, the terms ‘zostavax’ and ‘multiple people’ could now be part of the string representation because a larger number of smaller-sized clusters were allowed. Each instance of our hypothesis testing and corresponding results includes both these levels of granularity, adding another layer of robustness to our findings.

In this work, we combine all the texts we have: fake news articles, false fact-checked claims, and co-shared and control sets of mainstream news articles, and

this chapter, but for fake news articles containing many sentences per article, our work considers *recurring* narratives as *misinformation* narratives.

²¹See the code repository at <https://github.com/relatio-nlp/relatio>, which includes the default values used in this work.

Source of texts	Misinformation narratives library	Description	Total number of narratives (low-dimensional)	Total number of narratives (high-dimensional)
Fake news headlines & stories (30,120 texts)	<i>All fake-news narratives</i>	All narratives found in fake news texts	209,271	246,526
	<i>Recurring fake-news narratives</i>	Subset of above, top 1% in terms of frequency of occurrence	2,544	2,600
False fact-checked claims (21,109 texts)	<i>False-claims narratives</i>	Narratives found in claims that were fact-checked as false	10,838	11,384

Table 5.1: **Misinformation narrative libraries.**

apply the narrative extraction tool to this combined text data (ensuring that the various choices in narrative labels or string representations, such as the chosen entity representation for a cluster, remain consistent across all various texts). We then divide the output according to the different types of data being considered to get the set of narratives for each group of texts.

For misinformation narratives, we consider two sources of data: headline and story texts from the fake news articles ($\sim 24k$ URLs), and claims fact-checked to be false ($\sim 24k$ texts or claims). Headlines and story texts are used as separate text instances and processed to remove information like domain name mentions, which finally results in $\sim 30k$ texts.²² Misinformation narratives from fake news content are defined as recurring narratives: the top 1% in terms of frequency of occurrence. We also experiment with all the narratives found in fake news content (Table 5.1).

For fact-checked false claims, we collect and combine various publicly available datasets, and use the claims that are labeled as false (or an equivalent label such as ‘not true’; our codebase includes all the labels used to create our collection). These

²²The headlines are used as separate text instances since many (6-7 out of 10) people on social media only read the headline of an article (Gabiolkov et al., 2016); see also: <https://www.americanpressinstitute.org/publications/reports/survey-research/how-americans-get-news/>.

Narrative label	Sentences containing narrative label
baby survive abortion	1. the data on babies surviving abortions is compiled on a mandatory basis by only a handful of states. 2. should a baby survive an abortion, the health care practitioner shall "exercise the same degree of professional skill, care, and diligence to preserve the life and health of the child as a reasonably diligent and conscientious health care practitioner would render to any other child born alive at the same gestational age." 3. 77% of americans want to protect babies who survive abortions, nancy pelosi has blocked a vote 80 times 4. but data from minnesota, the centers for disease control and other sources indicate that babies do survive abortions, and laws are needed to protect them. 5. not only are women given the right to sue abortionists for the death of their babies that survive abortions, those in violation can be criminally charged in federal court for intentionally killing or attempting to kill a human being:
russian hack dnc	1. stone claims if the russians didn't hack the dnc, his case should be dismissed. 2. the mueller report clearly states that russians hacked the dnc and gave the hacked emails to wikileaks: 3. trending: the gateway pundit and american gulag donate \$40,000 to persecuted jan. 6 families this christmas on march 8, 2020 and before on june 16, 2019, we presented arguments against the mueller gang's assertion that the dnc was hacked by russians. 4. the news magazine claimed crowdstrike was correct in assessing that the dnc was hacked by russians in 2016. 5. when this news hit the media like a bombshell out-of-control left wing judge amy berman jackson tightened the unconstitutional gag order on roger stone to prevent him from discussing the shocking revelation and kravis filed a sur reply with the court falsely claiming that the us government and the mueller investigation had additional evidence to bolster their claim that the russians hacked the dnc.
bill stop infanticide	1. this is the second time senate democrats have blocked the bill to stop infanticide. 2. house democrats again block request to vote on bill to stop infanticide 3. nancy pelosi and democrats block bill to stop infanticide for 75th time, refuse care for babies born alive 4. democrats block bill to stop infanticide. 5. house democrats block request to vote on bill to stop infanticide

Table 5.2: Most frequent high-dimensional narrative strings or labels (and sentences containing those labels) in misinformation content, that are present in co-shared articles but not present in control articles published by mainstream 'real' news outlets.

claims are provided along with the rest of our data in our codebase.²³ Since these are statements, usually consisting of a single sentence fact-checked as false, we consider all narratives extracted from this data as misinformation narratives.

We thus create three misinformation narrative libraries, each containing low and high-dimensional narratives (Table 5.1). Examples of misinformation narratives are provided in Table 5.2.

5.4 Hypothesis Testing Procedure and Results

Our co-sharing method described above results in two groups of articles from reliable (“not-fake”) mainstream news media outlets: those disproportionately co-shared with at least one of the fake news articles, and those *not* co-shared with any of the fake news articles (the control group). We have also extracted and created different sets or libraries of misinformation narratives (Table 5.1).

Per our method, we also obtain both high- and low-dimensional narrative strings (or labels) from the headlines and story texts of co-shared and control mainstream articles. These enable us to test our hypothesis: **mainstream news stories that are disproportionately co-shared with fake news consist of a significantly higher presence of misinformation narratives**. Put another way, we test if misinformation narratives (extracted in various ways) are present in significantly higher proportions

²³Sources for the false fact-checked claims include *a*) claims fact-checked by FactCheck.org, Washington Post, and PolitiFact as collected by the Data Commons initiative (<https://www.datacommons.org/factcheck/download#research-data>); *b*) fact-checked public health claims Kotonya and Toni (2020) used to build automatic fact-checking systems that can also explain their classification; *c*) a benchmark dataset of fact-checked claims from PolitiFact (Wang, 2017); and *d*) a large-scale dataset of fact-checked claims coming from 26 different fact-checking websites (Augenstein et al., 2019). All claims are naturally occurring and in English. These publicly available claims dataset alongside fact-check labels are combined; this combined claims data along with the process of filtering for false claims is provided at https://github.com/Pranav-Goel/misinfo_narratives.

Type of outlets	Misinformation narrative library	Dimensionality of extracted narratives	n	
All reliable outlets	All fake news narratives	Low-dimensional	92577	$p < < < 0.05$
		High-dimensional	19845	$p < < < 0.05$
	Recurring fake news narratives	Low-dimensional	1982	$p < < < 0.05$
		High-dimensional	657	$p < < < 0.05$
	False-claims narratives	Low-dimensional	4413	$p < < 0.05$
		High-dimensional	318	$p < < < 0.05$
Only trustworthy outlets	All fake news narratives	Low-dimensional	86871	$p < < < 0.05$
		High-dimensional	15912	$p < < < 0.05$
	Recurring fake news narratives	Low-dimensional	1924	$p < < < 0.05$
		High-dimensional	587	$p < < < 0.05$
	False-claims narratives	Low-dimensional	4281	$p > 0.05$
		High-dimensional	306	$p < < 0.05$
Only liberal outlets	All fake news narratives	Low-dimensional	71654	$p < < < 0.05$
		High-dimensional	8235	$p < < < 0.05$
	Recurring fake news narratives	Low-dimensional	1856	$p < < < 0.05$
		High-dimensional	396	$p < < < 0.05$
	False-claims narratives	Low-dimensional	3956	$p < < < 0.05$
		High-dimensional	263	$p < < < 0.05$

Table 5.3: **Results** of significance testing showing that across various categories of reliable mainstream news outlets and sets of misinformation narratives, co-shared articles have a significantly higher presence of misinformation narratives compared with the control group.

in the co-shared set of articles than in the control group of articles published by the same mainstream outlets. If true, this would suggest that misinformation narratives may allow for a strategic repurposing of mainstream news in fake-news sharing circles on Twitter, something our data and method can be used to study as demonstrated by our case study (Section 5.5).

Our testing mechanism is visually explained in Fig. 5.1c. Consider a particular misinformation narrative library and a particular set of high- or low-dimensional narrative labels or strings in that library, and let's denote this particular collection as M , with size $|M|$ and narrative strings $n_1, \dots, n_i, \dots, n_{|M|}$. For both groups of mainstream articles (co-shared and control), we obtain the corresponding values of

each misinformation narrative as its normalized count in the set of all narratives extracted from that particular group of articles. Specifically, let C_s be the collection of all narrative strings present in co-shared mainstream articles (with $|C_s|$ total number of unique narrative strings: $s_1, \dots, s_i, \dots, s_{|C_s|}$), and C_o be the collection of all narrative strings present in control mainstream articles (with $|C_o|$ total number of unique strings: $o_1, \dots, o_i, \dots, o_{|C_o|}$). Then, the two lists of values we compare — l_{cs} for co-shared, and l_{co} for control — are created by iterating over all narratives (n_i) in M as follows:

$$l_{cs} = \left[\frac{\text{count}(n_i \in C_s)}{\sum_{i=1}^{|C_s|} \text{count}(s_i \in C_s)} \forall n_i \in M \right]$$

$$l_{co} = \left[\frac{\text{count}(n_i \in C_o)}{\sum_{i=1}^{|C_o|} \text{count}(o_i \in C_o)} \forall n_i \in M \right]$$

We then test if misinformation narratives have a significantly higher presence in co-shared articles ($l_{cs} > l_{co}$; Fig. 5.1c) using the one-sided Wilcoxon-signed rank test (non-parametric significance test for comparing paired or dependent data) (Wilcoxon, 1992; Woolson, 2007).²⁴

This process is repeated for all three of our misinformation narrative libraries (see Table 5.1), with the test performed separately for high- and low-dimensional narratives. To control for distributional shifts in the number of articles published by reliable domains, we also perform the same test with articles in both groups limited to those published by trustworthy outlets, as well as articles in both groups limited to those published by liberal outlets alone (trustworthy and liberal outlets are defined in the background section).²⁵ **The results across all these settings are**

²⁴Note that before this test is performed, any occurrences of ties and zeros were removed (Pratt, 1959), resulting in the final counts of the set of narratives being compared for their proportion or n shown in Table 5.3.

²⁵This check accounts for the fact that even when the two groups of mainstream articles are

	coef	std err	z	P >—z—
(Intercept)	0.4103	0.007	61.61	$< 2e - 16^{***}$
$X_{\text{misinfo_narrative}}$	0.2241	0.005	44.18	$< 2e - 16^{***}$

Table 5.4: Logistic regression results for $y \sim X_{\text{misinfo_narrative}}$, where $y = 1$ when the mainstream article is co-shared, and $y = 0$ when the mainstream article is control. $X_{\text{misinfo_narrative}}$ is the count for the number of narratives in the mainstream article that are also present in the misinformation narrative set. The results show that an article having a higher number of misinformation narratives is significantly and positively associated with co-sharing likelihood. The estimate can be roughly interpreted as: the co-shared group has 1.25 times the odds of the control group for containing a higher number of misinformation narratives. $***$ indicates significance at $p < 0.001$; and $1.25 \approx \exp^{0.2241}$

shown in Table 5.3 and clearly validate our hypothesis: articles from mainstream outlets that are co-shared with fake news contain misinformation narratives at a significantly higher rate.

To provide an estimate for the size of this effect, we also conduct a regression experiment, where every instance is a mainstream news article. The dependent variable (y) is a binary indicator for whether the mainstream article belongs to the co-shared or the control set. We combine all recurring narratives in fake news and all narratives in false fact-checked claims as our misinformation narrative set, and then count the number of narratives in the mainstream article that are also present in the misinformation narrative set. This count forms the independent variable (X or $X_{\text{misinfo_narrative}}$). Results in Table 5.4 show that there is a significant positive association between the number of misinformation narratives in an article and its co-sharing likelihood, with the co-shared group having roughly 1.25 times the odds of the control group for containing a higher number of misinformation narratives.

published by the same set of domains or outlets, the *number* of articles published by each domain can constitute different distributions. We show the distribution of per-outlet-article-counts for the two groups of articles from reliable outlets in Appendix D.1. It is clear when looking at the distribution of the number of articles per outlet that there is a conservative lean when considering all outlets or even just trustworthy outlets. This motivates performing our experiments with only liberal outlets as well.

Partisan curation of individual articles (measured using the ratio of registered Democrats and Republicans sharing the article on Twitter per [Green et al. \(2021\)](#)) is another viable explanation for the co-sharing of mainstream articles, and indeed, we find it also represents a significant difference between co-shared and control articles. Our additional experiments testing the robustness of misinformation narratives as an explanation for the co-sharing phenomenon in the presence of partisan curation are detailed in Appendix [D.2](#).²⁶ We find that partisanship and misinformation narratives are two distinct, significant, and complementary factors when explaining the difference between co-shared and control sets of mainstream articles.

5.5 Case Study

Our quantitative results show that narratives constitute a statistically significant explanation or factor in the co-sharing of certain mainstream articles in the *aggregate*. In addition, our data and methods can help flesh out how narratives can utilize a mix of fake and ‘real’ news, and why and how particular mainstream articles co-shared with fake news can help support misleading narratives. To this end, we now highlight how our work can contribute to the understanding of specific narratives in the media landscape, and *enable* journalists and scholars to help investigate potential cases of synergistic support of misinformation and misleading narratives in mainstream news articles via a comprehensive case study.

In our manual examination of narrative labels or strings (and narratives) present in co-shared articles published by liberal mainstream outlets, we find multiple labels of the form ‘vaccinated people spread variant, disease, or virus’.

²⁶Results corresponding to the regression results shown in Table [5.4](#) but with another independent variable for the partisanship of the article are provided in Table [D.2](#); the odds for the co-shared group compared with control group remain around 1.2 in the presence of a variable for partisanship.

In this case study:

- We first detail our manual filtering approach to form a set of such aforementioned narrative labels and present the selected narrative labels along with the sentences containing those labels (in order to facilitate the interpretation of the narratives themselves).
- We then provide and examine all the mainstream articles that are published by liberal news outlets and co-shared with fake news that contain these narratives. We identify and qualitatively divide these mainstream articles into three types of articles (in terms of their propensity to be used for misleading purposes). We discuss the types likely to be co-opted for misleading purposes in detail.
- To understand the world of fake news these co-shared mainstream articles are interacting with, we show a sample of the vaccine-related fake news articles the aforementioned mainstream articles are co-shared with.
- Crucially, to better understand *how* the aforementioned interactions are occurring, we show a sample of tweets authored by conveyors of vaccine-related fake news that shared the identified co-shared articles from liberal-leaning mainstream outlets. This can provide a window into the way some of the ‘real’ news may be getting used to support misinformation or misleading agendas.
- Finally, we quantify the potential impact of one particular type of co-shared mainstream news: when the headline is framed in a somewhat clickbait manner, which could allow for relatively straightforward repurposing in order to push anti-vaccination misinformation narratives.

Narrative Label	Sentences containing narrative
	1. though biden has received both shots of the pfizer coronavirus vaccine, centers for disease control and prevention guidelines stipulate that vaccinated people are to continue wearing face masks, citing a lack of research showing whether those vaccinated can still spread the virus. 2. the killed vaccine doesn't fully block the virus from spreading because a person who is immunized can still carry and spread the polio virus. 3. first, can someone who has been vaccinated still spread the disease? 4. covid has provided a grim illustration of how la's longstanding housing crisis is a public health catastrophe; the region is a leader in overcrowded housing, which has further spread the virus among vulnerable populations. 5. in yet another unexpected and unwelcome twist in the pandemic, the centers for disease control and prevention released on friday a report strongly suggesting that fully immunized people with so-called breakthrough infections of the delta variant can spread the virus to others just as readily as unvaccinated people.
'vaccinate people spread delta variant'	1. wakana addressed the same issue with a new york times tweet stating "breaking news: the delta variant is as contagious as chickenpox and may be spread by vaccinated people as easily as the unvaccinated, an internal c.d.c.
'fully vaccinate people spread delta variant'	2. top health expert says vaccinated people are spreading delta variant 1. cdc says fully vaccinated people spread the delta variant and should wear masks: 'this new science is worrisome'

Table 5.5: Part one of the full set of selected narrative labels capturing the type of narrative we examine in our case study.

Narrative Label	Sentences containing narrative
'vaccinate people spread virus'	1. at no point did the vaccine promise absolute protection from transmitting the disease anyway, so it always felt like the most likely outcome of masklessness was that unvaccinated people would spread their germs to vaccinated people , who then, even if they didn't get sick, would spread the virus back to unvaccinated people , lengthening the time everybody has to spend in this hellish cycle.
	2. but scientists do not yet know if vaccinated people could still spread the virus .
	3. vaccinated people may spread the virus , though rarely, c.d.c.
	4. and vaccinated people can still spread the virus .
	5. plus, while vaccinated people are much less likely to get infected, and therefore much less likely to spread the virus , it's still not entirely clear that they are not contagious, if they do happen to get infected.
	6. by comparing infection rates among the contacts of those who were vaccinated and those who weren't, scientists can get a better idea of whether vaccinated people can spread the virus .
'fully vaccinate individual spread virus'	1. the agency did not rule out the possibility that fully vaccinated individuals might develop asymptomatic infections and spread the virus inadvertently to others, and urged those who are vaccinated to continue practicing certain precautions.
'vaccinate people infect delta variant vaccination spread virus'	1. "information on the delta variant from several states and other countries indicates that, on rare occasions, some vaccinated people infected with the delta variant after vaccination may be contagious and spread the virus to others," walensky said during a press call on tuesday.
'vaccinate people spread disease'	1. while the data suggests that vaccinated people can spread the disease , the extent to which they contribute is not yet clear.
'data suggests vaccinate people spread disease'	1. while the data suggests that vaccinated people can spread the disease , the extent to which they contribute is not yet clear.

Table 5.6: Part two of the full set of selected narrative labels capturing the type of narrative we examine in our case study.

Narrative Label	Sentences containing narrative
'vaccinate spread coronavirus'	1. but ultimately, a covid-19 diagnosis in someone close to me is what highlighted why the assertion that the vaccinated are as likely to spread the coronavirus as the unvaccinated is so wrong. 2. using tear gas to subdue protesters may further spread the coronavirus, experts warn
'vaccinate people spread coronavirus'	1. among this last group, a common refrain i've heard to justify their renewed vigilance is that "vaccinated people are just as likely to spread the coronavirus." 2. so let me make one thing clear: vaccinated people are not as likely to spread the coronavirus as the unvaccinated. 3. but they could also be a silent spreader," he said, referring to the fact that it has not yet been determined if vaccinated people can still carry and spread the coronavirus.
'scientist not-ignore vaccinate people spread virus'	1. but scientists do not yet know if vaccinated people could still spread the virus.
'vaccine allows disease spread faster longer normally'	1. in fact, rather than stop fowl from spreading the virus, the vaccine allows the disease to spread faster and longer than it normally would, a new study finds.

Table 5.7: Final part (part three) of the full set of selected narrative labels capturing the type of narrative we examine in our case study.

Selecting narratives using the narrative labels. To study the narratives that might be attributing the spread of disease or virus to vaccines or vaccinated people, we first obtain all narrative strings or labels in co-shared articles by liberal mainstream outlets that contain the strings ‘vaccin’ and ‘spread’. The full set of labels obtained after this step is provided in Appendix [D.3](#).

We then manually filter them down to coherent narrative labels fitting the type of narrative we are investigating in this case study. These labels (such as ‘vaccinate people spread virus’) along with the sentences containing them are provided in Tables [5.5](#) to [5.7](#) and help understand the type of narrative that forms the subject of this case study.

Examining and classifying mainstream articles published by liberal outlets containing selected narratives. All articles published by liberal news outlets that are co-shared with fake news and contain narratives displayed in Tables [5.5](#) to [5.7](#) are provided in Tables [5.8](#) to [5.10](#). These nine articles, in terms of their propensity to be used for misleading purposes, can be grouped qualitatively into three types:

- **Type 1: Headlines that are somewhat clickbait.** We find that three out of the nine mainstream articles seem to contain at least a mild form of clickbait or misleading headlines, via trivialization or simplification that could provide easy support for anti-vaccination misinformation narratives and beliefs.

Headlines carry particular significance in their much wider reach and readership ([Dor, 2003](#); [Gabiolkov et al., 2016](#)) as well as influence on how the article, if read, is interpreted ([Otero and Kintsch, 1992](#); [McCrudden and Schraw, 2007](#); [Ecker et al., 2014](#)). When written for the primary purpose of generating clicks rather than to inform, headlines can eschew contexts in ways that can sensationalize via presenting the news item as “more interesting, extraordinary and

Type	Headline (URL)	Article Text Snippet	Narrative Labels Contained
Type 1	Top health expert says vaccinated people are spreading delta variant https://thehill.com/changing-america/well-being/longevity/561994-top-health-expert-says-vaccinated-people-are-spreading	As some states report marked upticks in COVID-19 infections while the delta variant becomes the most common strain in the U.S., public health experts are encouraging further vaccinations to help curb transmission. Some experts, however, warn that vaccinated individuals may still be capable of contracting and transmitting COVID-19. ... Other data emerging from Israel suggest that the Pfizer vaccine is less effective against infections caused by the delta variant, which may explain the uptick in U.S. infection despite widespread vaccinations.	[vaccinate people spread delta variant]
Type 1	CDC says fully vaccinated people spread the Delta variant and should wear masks: https://www.businessinsider.com/cdc-fully-vaccinated-people-can-spread-delta-variant-2021-7	The CDC updated its guidelines on Tuesday to recommend masks indoors, even for vaccinated people. The Delta variant makes it easier for vaccinated people to transmit the virus, the CDC said. But vaccines remain highly effective at preventing severe disease, no matter the variant. ... The goal of these new guidelines, CDC director Rochelle Walensky said, is to “help prevent the spread of the Delta variant and protect others.” “Information on the Delta variant from several states and other countries indicates that, on rare occasions, some vaccinated people infected with the Delta variant after vaccination may be contagious and spread the virus to others,” Walensky said during a press call on Tuesday. “This new science is worrisome and unfortunately warrants an update to our recommendations.” Walensky said CDC investigations have found that the amount of virus present in vaccinated people infected with Delta is similar to the levels found in unvaccinated people with Delta infections. That’s an indication that vaccinated people can easily transmit the virus — even though they’re less likely to get sick on the whole. ...	[fully vaccinate people spread delta variant, vaccinate people infect delta variant vaccination spread virus]
Type 1	CDC study shows three-fourths of people infected in Massachusetts coronavirus outbreak were vaccinated but few required hospitalization https://www.washingtonpost.com/health/2021/07/30/provincetown-covid-outbreak-vaccinated/	... Scientists said the Provincetown outbreak and other recent data on breakthrough infections make clear that vaccines offer significant protection, as they were designed to, against severe illness and death but do not offer blanket protection against any chance of infection. Only a handful of people in the outbreak were hospitalized. While the data suggests that vaccinated people can spread the disease, the extent to which they contribute is not yet clear. Walensky said this week that such transmission occurs on “rare occasions.” ...	[vaccinate people spread disease, data suggests vaccinate people spread disease]

Table 5.8: Part one of all the articles (qualitatively categorized as type 1) published by liberal outlets that are co-shared with fake news and contain selected narratives for our case study examining a particular type of narrative of the form ‘vaccinated people spread (virus, disease, variant, etc.).’

Type	Headline (URL)	Article Text Snippet	Narrative Labels Contained
Type 2	Mutant Strains Of Polio Vaccine Now Cause More Paralysis Than Wild Polio (https://www.npr.org/sections/goatsandsoda/2017/06/28/534403083/mutant-strains-of-polio-vaccine-now-cause-more-paralysis-than-wild-polio) (published in June 2017)	For the first time, the number of children paralyzed by mutant strains of the polio vaccine are greater than the number of children paralyzed by polio itself. ... And finally, the live vaccine stops transmission of the polio virus entirely in a community if sufficient numbers of people are vaccinated. The killed vaccine doesn't fully block the virus from spreading because a person who is immunized can still carry and spread the polio virus . And this is an important difference between these two types of vaccines when the goal is to exterminate the polio virus. "The fact is this [the live oral polio vaccine] is the only tool that we have that can eradicate the disease," says Zaffran. ... Zaffran credits the oral polio vaccine with getting the world incredibly close to wiping out a terrible disease. ...	[vaccinate spread virus]
Type 2	This chicken vaccine makes its virus more dangerous (https://www.pbs.org/newshour/science/chicken-vaccine-makes-virus-dangerous) (published in July 2015)	... Chickens vaccinated against Marek's disease rarely get sick. But the vaccine does not prevent them from spreading Marek's to unvaccinated birds. "With the hottest strains, every unvaccinated bird dies within 10 days. There is no human virus that is that hot. Ebola, for example, doesn't kill everything in 10 days." In fact, rather than stop fowl from spreading the virus, the vaccine allows the disease to spread faster and longer than it normally would, a new study finds. ...	[vaccine allows disease spread faster longer normally]

Table 5.9: Part two of all the articles (qualitatively categorized as type 2) published by liberal outlets that are co-shared with fake news and contain selected narratives for our case study examining a particular type of narrative of the form 'vaccinated people spread (virus, disease, variant, etc.).'

Type	Headline (URL)	Article Text Snippet	Narrative Labels Contained
Type 3	The Delta Variant Is Contagious as Hell—and People Are Selfish (https://www.thenation.com/article/society/covid-delta-kids)	... On Thursday, we learned the reason for the CDC's latest flip-flop in mask policy. An internal memo, obtained by The Washington Post, cautioned that "the war has changed," and outlined the various ways the Delta variant of Covid-19 is more contagious as well as the likelihood that it can be spread by people who have already been vaccinated At no point did the vaccine promise absolute protection from transmitting the disease anyway, so it always felt like the most likely outcome of masklessness was that unvaccinated people would spread their germs to vaccinated people , who then, even if they didn't get sick, would spread the virus back to unvaccinated people, lengthening the time everybody has to spend in this hellish cycle. ...	[vaccinate people spread virus]
Type 3	How 700 Epidemiologists Are Living Now, and What They Think Is Next (https://www.nytimes.com/2020/12/04/upshot/epidemiologists-virus-survey.html)	... Moncef Slaoui, who is leading the government's Operation Warp Speed vaccine development program, said this week that vaccines may roll out quickly enough for the United States to reach herd immunity by early summer. But scientists do not yet know if vaccinated people could still spread the virus	[vaccinate people spread virus, scientist not ignore vaccinate people spread virus]
Type 3	Vaccinated Americans may gather indoors in small groups but should still wear masks in public, the C.D.C. said. (https://www.nytimes.com/2021/03/08/health/covid-vaccine-cdc-guidelines.html)	... The agency did not rule out the possibility that fully vaccinated individuals might develop asymptomatic infections and spread the virus inadvertently to others, and urged those who are vaccinated to continue practicing certain precautions. ...	[fully vaccinate individual spread virus]
Type 3	No, Vaccinated People Are Not 'Just as Likely' to Spread the Coronavirus as Unvaccinated People (https://www.theatlantic.com/ideas/archive/2021/09/the-vaccinated-arent-just-as-likely-to-spread-covid/620161)	... Many are still taking no chances. Some have even returned to all the same precautions they took before getting their shots, including avoiding the company of other fully vaccinated people. Among this last group, a common refrain I've heard to justify their renewed vigilance is that " vaccinated people are just as likely to spread the coronavirus ." ... So let me make one thing clear: Vaccinated people are not as likely to spread the coronavirus as the unvaccinated. ... Relax, America: the vaccines are still working. But ultimately, a COVID-19 diagnosis in someone close to me is what highlighted why the assertion that the vaccinated are as likely to spread the coronavirus as the unvaccinated is so wrong. ...	[vaccinate spread coronavirus, vaccinate people spread coronavirus]

Table 5.10: The final part (part three) of all the articles (qualitatively categorized as type 3) published by liberal outlets that are co-shared with fake news and contain selected narratives for our case study examining a particular type of narrative of the form 'vaccinated people spread (virus, disease, variant, etc.).'

relevant than might be the case” (Molek-Kozakowska, 2013). Clickbait headlines that implement sensationalism have been studied for their propensity to mislead (Chen et al., 2015; Blom and Hansen, 2015; Biyani et al., 2016).²⁷ A recent study by Gabriel et al. (2022) studies the pragmatic implications of news headlines and the effect news headlines have on readers, and introduces the Misinfo Reaction Frames (MRF) corpus along with NLP models trained on that corpus. Their findings suggest that what news headlines imply in terms of their pragmatics influences the readers’ reactions, and reasoning about the implications of headlines is important for combating misinformation.

In these types of articles that emerge in our case study, for example, a headline such as “Top health expert says vaccinated people are spreading delta variant” (first row in Table 5.8) can be used to ascribe a causal or unique role to vaccinated people in the spread of the variant as opposed to unvaccinated people (not implicated via omission). Compare the headline with this piece of writing in the article: “...public health experts are encouraging further vaccinations to help curb transmission. Some experts, *however*, warn that vaccinated individuals *may still* be capable of contracting and transmitting COVID-19.” (emphasis is ours). The addition of uncertainty (in direct contrast to the headline) as well as the clarity that rather than a unique or causal role, it is the fact that vaccines may not put a complete stop to transmissions, becomes apparent when reading the article. In this manner, another form of misleadingness present in these types of articles might be the phenomenon of incongruent headlines (Chesney et al., 2017) where the headlines are not necessarily entailed by the contents of the article.

²⁷Note that while we use clickbait headlines with a somewhat negative connotation given their propensity to mislead, some prior work has also examined the potentially positive role of clickbait headlines by attracting readers to reliable and important news reports (Bhowmik et al., 2019).

The repurposing of such content by vaccine misinformation conveyors on Twitter is further examined below by looking at the texts of the tweets sharing these articles.

- **Type 2: Older news articles in potentially new contexts.** These are articles from before the period of our Twitter dataset tweets reporting on particular vaccine failures of the past. As shown in Table 5.9, we find two such articles, one from mid-2017 and another from 2015 (whereas our Twitter dataset consists of tweets from May 2018 to November 2021). These articles are reporting on the failures of polio and chicken vaccines. In the time period covered by our dataset, COVID-19 was the most salient focus of vaccination discussion.

Sharing older news articles in potentially a new context can be used to mislead as noted in Rogers (2019). In fact, Rogers (2019) noted sharing of older mainstream news articles as a disinformation tactic and also reported on the steps being taken by news organizations to highlight their publication date. In 2020, Facebook rolled out a feature to notify users when they are sharing articles more than 90 days old.²⁸ The tweets by vaccine misinformation conveyors sharing these type-2 liberal co-shared articles are discussed below and help understand this kind of temporal repurposing of legitimate reporting.

- **Type 3: Other.** These are simply the remaining articles, and they do not seem to be easily usable for misleading purposes.

Sample of fake news articles co-shared with identified liberal outlet articles. For the mainstream articles presented in Tables 5.8 to 5.10, we obtain all the fake news

²⁸See <https://www.dailymail.co.uk/sciencetech/article-8460273/Facebook-tell-users-share-old-news-article.html> and <https://www.forbes.com/sites/alisondurkee/2020/06/26/new-facebook-policy-warns-users-before-they-share-old-articles/?sh=3b0eb9fb3f34>

Co-Shared Vaccine-related Fake News URL	Co-Shared Vaccine-related Fake News Headline	Corresponding Co-shared Liberal "Real" News URL
https://childrenshealthdefense.org/news/government-corruption/gates-globalist-vaccine-agenda-a-win-win-for-pharma-and-mandatory-vaccination	Gates' Globalist Vaccine Agenda: A Win-Win for Pharma and Mandatory Vaccination	[https://www.npr.org/sections/goatsandsoda/2017/06/28/534403083/mutant-strains-of-polio-vaccine-now-cause-more-paralysis-than-wild-polio , https://www.washingtonpost.com/health/2021/07/30/provincetown-covid-outbreak-vaccinated/]
https://www.lifesitenews.com/news/ontario-er-doctor-resigns-over-mandatory-vaccines-and-falsehoods	Ontario doctor resigns over forced vaccines, says 80% of ER patients with mysterious issues had both shots	[https://thehill.com/changing-america/well-being/longevity/561994-top-health-expert-says-vaccinated-people-are-spreading , https://www.washingtonpost.com/health/2021/07/30/provincetown-covid-outbreak-vaccinated/]
https://childrenshealthdefense.org/defender/americas-frontline-doctors-federal-lawsuit-halt-covid-vaccines-cdc-vaccine-deaths	Federal Lawsuit Seeks Immediate Halt of COVID Vaccines, Cites Whistleblower Testimony Claiming CDC Is Under-Counting Vaccine Deaths	[https://thehill.com/changing-america/well-being/longevity/561994-top-health-expert-says-vaccinated-people-are-spreading , https://www.washingtonpost.com/health/2021/07/30/provincetown-covid-outbreak-vaccinated/]
https://thefederalist.com/2021/10/07/joe-bidens-vaccine-mandate-doesnt-exist-its-just-a-press-release	Joe Biden's Vaccine Mandate Doesn't Exist. It's Just A Press Release	[https://www.washingtonpost.com/health/2021/07/30/provincetown-covid-outbreak-vaccinated , https://www.pbs.org/newshour/science/tthis-chicken-vaccine-makes-virus-dangerous]
https://healthimpactnews.com/2014/guillain-barre-syndrome-is-1-side-effect-of-vaccine-injury-compensations-due-to-flu-shots	Guillain Barré Syndrome is #1 Side Effect of Vaccine Injury Compensations due to Flu Shots	[https://www.npr.org/sections/goatsandsoda/2017/06/28/534403083/mutant-strains-of-polio-vaccine-now-cause-more-paralysis-than-wild-polio]
https://vactruth.com/vaccine-inserts-human	Human Vaccine Inserts	[https://www.npr.org/sections/goatsandsoda/2017/06/28/534403083/mutant-strains-of-polio-vaccine-now-cause-more-paralysis-than-wild-polio]
https://healthimpactnews.com/2017/harvard-immunologist-to-legislators-unvaccinated-children-pose-zero-risk-to-anyone	Harvard Immunologist to Legislators: Unvaccinated Children Pose ZERO Risk to Anyone	[https://www.npr.org/sections/goatsandsoda/2017/06/28/534403083/mutant-strains-of-polio-vaccine-now-cause-more-paralysis-than-wild-polio]
https://thehighwire.com/study-cdc-vaccine-schedule-likely-induces-aluminum-toxicity-in-newborns	STUDY: CDC Vaccine Schedule Likely Induces Aluminum Toxicity in Newborns	[https://www.npr.org/sections/goatsandsoda/2017/06/28/534403083/mutant-strains-of-polio-vaccine-now-cause-more-paralysis-than-wild-polio]

Table 5.11: Vaccine-related fake news significantly co-shared with identified liberal mainstream articles in our case study.

articles that are (significantly) co-shared with one or more of these ‘real’ articles and contain the string ‘vaccin’ in their (lowercased) headline (to make it more likely that the fake news article is about vaccines or vaccination). A total of 61 such co-shared vaccine-related fake news articles were found in our dataset. We show all four fake news articles that were co-shared with more than one of the identified liberal news articles, and a random sample of four more fake news articles that were co-shared with one of the identified liberal news articles in Table 5.11. This helps illustrate the vaccine-related fake news content the identified mainstream news articles are interacting with on Twitter and helps emphasize why we are studying the misinformation-adjacent mainstream news in this chapter.

Examining *how* mainstream news articles might be getting used for misleading purposes by considering the tweet texts. For type 1 (articles with somewhat clickbait-y headlines) and type 2 (older articles potentially shared in new contexts) liberal outlet articles identified above, we look at tweet texts sharing those particular articles by conveyors of vaccine-related fake news. This helps us understand how these articles might be getting used to support misleading narratives.

We first consider all fake news articles in our dataset for which we have automatically extracted headlines. We obtain all fake news articles with the string ‘vaccin’ in their (lowercased) headline, which results in 429 vaccine-related fake news. We filter our entire Twitter panel data to tweets by users who shared vaccine-related fake news five or more times, resulting in 1074 different Twitter users. For each type 1 and type 2 mainstream article, we obtain all unique tweet texts that share those articles by this subset of vaccine-misinformation-sharing Twitter users.²⁹

A sample of up to five tweet texts is provided alongside the mainstream article

²⁹All username mentions are removed from the tweet texts and only texts with at least three space-separated tokens are retained.

URL	Tweet texts
https://thehill.com/changing-america/well-being/longevity/561994-top-health-expert-says-vaccinated-people-are-spreading	[Top health expert says vaccinated people are spreading delta variant — TheHill, Vaccines that target the spike protein are causing it to mutate. Virology 101. Top health expert says vaccinated people are spreading delta variant — TheHill, Top health expert says vaccinated people are spreading delta variant TheHill - THAT IS LITERALLY THE HEADLINE. I wonder if President Biden will consider me a “murderer” for sharing., Top health expert says vaccinated people are spreading delta variant]
https://www.businessinsider.com/cdc-fully-vaccinated-people-can-spread-delta-variant-2021-7	[CDC says fully vaccinated people spread the Delta variant and should wear masks: ‘This new science is worrisome’ ALL YOUR “SCIENCE” IS WORRISOME!!., BuT hoW ManY maskS sHOuld thEy WeAr?? CDC says fully vaccinated people spread the Delta variant and should wear masks: ‘This new science is worrisome’ via @businessinsider, What will tomorrow bring? A new set of opinions? CDC says fully vaccinated people spread the Delta variant and should wear masks: ‘This new science is worrisome’ via @businessinsider, Vaccinated and unvaccinated carry same viral loads, per the CDC.. KEEP THIS SHOW IT TO YOUR STUPID VACCINATED FRIENDS AND FAMILY, YOUR GOING TO NEED IT., Well look what the CDC is saying today It’s the vaccinated spreading covid.]
https://www.washingtonpost.com/health/2021/07/30/provincetown-covid-outbreak-vaccinated	[Good grief. Did not beat Covid variant. Mask up again. Booster shots? CDC study shows three-fourths of people infected in Massachusetts covid-19 outbreak were vaccinated, ¾ were vaxed in MA outbreak, Vaccinated people made up three-quarters of those infected in a massive Massachusetts covid-19 outbreak, pivotal CDC study finds, Zero science to back up any of this. Even CDC says vaccinated still spread #COVID; mask mandated never reduced or prevented spread of #covid, The proper response is not to blame Blue America for outbreak but for those finger-pointing commentators to be a little more circumspect, a little less Trump-obsessed, a little more reasonable. Would be good idea. 3/3 From WP.]
https://www.npr.org/sections/goatsandsoda/2017/06/28/534403083/mutant-strains-of-polio-vaccine-now-cause-more-paralysis-than-wild-polio	[The oral polio vaccine spreads polio. Every case is caused by the vaccine. No child needs 72 doses of vaccines. Hep B vax given on day 1 of life is insane., 2 year old article is still relevant., Nope. Vaccines have been causing VARIANTS for a LONG time. All you folks that got VACCINATED are the ones who have “caused this”. #TheVaccinesAreTheVariants, None for me, thanks. I’ll take my polio natural., Mutant Strains of #Polio #Vaccine Now Cause More Paralysis Than Wild Polio 2017: for the first time, the # of children paralyzed by mutant strains of the polio vaccine are greater than the number of children paralyzed by polio itself. #vaccines #vexit]
https://www.pbs.org/newshour/science/tthis-chicken-vaccine-makes-virus-dangerous	[Wait until he hears about (see latest tweets) and Marek’s Chickens and finds out the vaxxed are the real superspreaders., Just realize: 1) We’ve never mass vaccinated in middle of a circulating pandemic 2)We’ve never used such a narrow spectrum spike protein 3) We’ve never used such a leaky vaccine in terms of efficacy against transmission With that in mind, now read this, We’ve known since at least 2015 that an #ImperfectVaccine can make a virus worse. Repeating the same mistake is insanity., This chicken vaccine makes its virus more dangerous — PBS NewsHour, For those interested, here’s a short review.]

Table 5.12: Sample of tweets texts when vaccine-misinformation-conveyors on Twitter share particular mainstream articles identified in our case study.

in Table 5.12, and highlights how these various articles can be used to buttress anti-vaccine misinformation narratives. We also present all the unique tweet texts from which the tweets presented in Table 5.12 are randomly sampled, in Appendix D.4.

Impact of framing choices in mainstream articles: disproportionate uptake amongst conveyors of misinformation. We take a closer look at the three type-1 mainstream articles (Table 5.8), where the framing choices in the headlines seem to facilitate repurposing those articles towards the push for certain misleading narratives by conveyors of vaccine misinformation on Twitter (Table 5.12). A consequence of those framing choices is those articles being co-shared with fake news. We further find that those particular articles by the three mainstream outlets (The Hill, Business Insider, and The Washington Post) are some of their most popular vaccine-related articles among vaccine-misinformation-sharers relative to all the vaccine reporting done by those outlets (per the URLs shared in our Twitter panel between May 2018 and November 2021). For example, out of all the 444 vaccine-related articles published by The Washington Post in our final dataset, the particular co-shared URL (<https://www.washingtonpost.com/health/2021/07/30/provincetown-covid-outbreak-vaccinated>) with the headline ‘CDC study shows three-fourths of people infected in Massachusetts coronavirus outbreak were vaccinated but few required hospitalization’ was the 2nd highest in terms of shares amongst the subset of Twitter users known to share vaccine-related fake news (identified using the procedure discussed above).

We also create a subset of Twitter users in our dataset that *a*) did not share any fake news article in our entire dataset, and *b*) shared vaccine-related articles from liberal mainstream outlets at least five times. For all three articles, we compare the popularity (in terms of number of shares) of the article amongst the vaccine fake news sharers and vaccine ‘real’ news sharers, relative to all vaccine-related articles published by that corresponding outlet. We find that these articles got significantly higher attention (in terms of Twitter sharing) amongst vaccine-related fake news conveyors, providing a clear measure of the impact of the framing choices

made in the mainstream articles and the disproportionate uptake in misinformation circles on Twitter. Details of our procedure and significance testing are provided in Appendix [D.5](#).

5.6 Discussion and Related Work

Consistent with our hypothesis, we find that a significantly higher presence of misinformation narratives is associated with disproportionate uptake of mainstream news articles by conveyors of fake news on Twitter. This effect — significantly higher misinformation narratives in mainstream news that are co-shared with fake news — is not explained away by partisan curation, *i.e.*, it is not just a case of news of a particular partisan leaning being popular in fake-news-circles on Twitter when it comes to misinformation narratives pulling together different kinds of sources. Our finding suggests that mainstream news is part of how misinformation narratives are pushed on social media and that it might be used to enhance the reach and persuasiveness of misleading narratives.

To the extent that this hypothesis is true, it has important implications for responsible journalistic practice: when vetting news stories and their framing, especially for the headlines, it is important to consider not just the story itself but also the broader context of existing narratives in fake news. Technological methods like the ones we have developed and used here may help go beyond checking a story’s factuality and accuracy and contribute to minimizing the risk of a story containing misinformation narratives and being co-opted by conveyors of fake news on social media.

For scientific research on online misinformation, our work highlights the importance of looking at not just falsehoods in fake news, but the larger narratives

and world views being fleshed out — and examining the role of communicative choices (in the form of existing misinformation narratives) made in mainstream news (that would generally not be deemed fake or factually incorrect). Indeed, a recent study posits that focusing on misinformation leads to misunderstanding the presence of misleading information and widespread occurrence of misperceptions despite the fact that fake news itself, as we mentioned earlier, is a very tiny part of people’s media consumption (Williams, 2022). Instead, Williams suggests that a better framework for understanding the apparent problems with the online information ecosystem is the synthesis of motivated reasoning and the (new) concept of *a marketplace of rationalizations*. In brief, sometimes, humans are motivated to believe things out of various forms of desirability (due to emotional, social, or material benefit) rather than accuracy, but there is also a constraint to present the belief in a manner that appears objective and *rational*. Authors and institutions can serve to offer rationalizations for a variety of beliefs and worldviews (constituting a marketplace). Fake news can be seen as part of such material that provides rationalizations for some pre-existing beliefs.

Building upon this theory, we consider it plausible that the narratives present in fake news content (such as “vaccines harm people”) can signal the world views and beliefs for which that particular piece of content provides rationalizations. Articles by mainstream outlets that are co-shared with fake news (such as the WAPo article highlighted in Fig. 5.1 originally titled “Vaccinated people now make up a majority of covid deaths”) might become a part of the rationalization process for misperceptions and misleading narratives or beliefs. Our case study in Section 5.5 suggests as much, with tweets such as “Well look what the CDC is saying today It’s the vaccinated spreading covid.” sharing the article <https://www.businessinsider.com/cdc-fully-vaccinated-people-can-spread-delta-variant-2021-7>

with the headline “CDC says fully vaccinated people spread the Delta variant and should wear masks: ‘This new science is worrisome’.” The same misleading narrative can now be rationalized using a combination of different information sources (fake news, mainstream news, cross-partisan outlets). This accords with our finding that the presence of existing misinformation narratives characterizes the particular articles from mainstream outlets co-shared with fake news.

Very few users share articles from known fake news outlets ([Grinberg et al., 2019](#); [Altay et al., 2022a](#)). [Altay et al. \(2022a\)](#) suggest that one reason for that is the potential harm to online social reputation as a result of sharing content from outlets not granted mainstream legitimacy. Sharing mainstream content containing misleading narratives and beliefs present in fake news could provide a way to circumvent those potential reputational harms while still sharing those narratives.

The presence of existing misinformation narratives in the mainstream media landscape also hints at the potential agenda-setting power of fake news. For example, [Vargo et al. \(2018\)](#) studied how fake news content both responds to and influences the agenda in more mainstream partisan media, and find that especially in 2016, fake news may have driven the agenda for partisan media in the US, while the findings in [Tsfati et al. \(2020\)](#) suggest that mainstream news outlets are sometimes responsible for disseminating and popularizing particular fake news stories (even more than social media). [Guo and Vargo \(2020\)](#) then further studied *how* the interaction between fake news outlets and mainstream media takes place. Their findings suggest that fake news outlets might be utilizing the sensationalism *already* present in the mainstream media environment and “add noise” to that, instead of directly setting the agenda for news outlets. This latter study, concordant with our work, once again confirms the responsibility and need for structural action in the mainstream journalism landscape itself rather than confining the problem of

misinformation to a certain set of identified ‘bad actors’ in the form of specific fake news outlets.

Related to our work and findings, recent work in [Sharma et al. \(2022\)](#) investigated COVID-19 vaccination misinformation narratives on social media (whereas we look at the content of news articles).³⁰ The narratives they found in social media communities are similar to ones we extracted from fake news content and fact-checked false claims (which do include claims from social media): narratives questioning vaccine safety and effectiveness. Interestingly, [Sharma et al. \(2022\)](#) noted the “distortion of facts due to misleading rather than outright false narratives” such as exaggeration of side-effects and recontextualization. This mirrors our findings in terms of observed extracted narratives, the co-shared fake and mainstream news content, and the indication of strategic repurposing of mainstream news content on social media (which can help support efforts to recontextualize legitimate reporting). The latter indication of repurposing and possibly taking advantage of how headlines and mainstream story content are framed is backed by the findings of our case study (Section 5.5).

5.7 Limitations and Future Work

For our quantitative results in this study, we examine the content of articles, but for tweets, we only considered the articles shared and not the text generated by the user. Our findings, the aforementioned concept of rationalizations, and our case study all suggest that future work should also look at the *content of tweets* sharing news articles, robustly identify misinformation sharers, and study *how* mainstream news articles get repurposed to support pre-existing misperceptions

³⁰We note that narratives are conceived more broadly than sets of entity-action relations by [Sharma et al. \(2022\)](#).

via the potential choices made in tweet texts (while maintaining an understanding of the larger media ecosystem that enables such repurposing). Examining tweets can also help verify what our findings indicate: mainstream outlets might be used to grant an air of legitimacy to certain narratives. Such an examination can help analyze the instances where mainstream news content itself misleads *directly* and instances where mainstream news tends to *enable* misleading usage by users on Twitter.

Another subject of future analysis would be examining where and how misinformation narratives occur in mainstream news stories. One possibility is that these narratives do not occur in the content written by the journalist or news reporter themselves, but in *direct quotations*: content attributed to a different actor or source replicated and used verbatim in a text. Such direct quotations are marked clearly using quotation markers (such as "...") in texts (Pouliquen et al., 2007). We aggressively removed instances of direct quotations from co-shared and control articles in a preliminary analysis (aiming for recall), and find that our hypothesis holds when working with this modified content. Therefore, the significantly higher presence of these misinformation narratives in co-shared articles does not seem to be driven by the practice of direct quotations in media alone. Overall, our effect holds for these newly processed texts with direct quotations removed in 16/18 significance testing instances corresponding to rows in Table 5.3. We report our preliminary analysis including the procedure to remove direct quotations and detailed results in Appendix D.6 to help inform future work.

In any case, future work should examine if the practice of direct quotations (and the possible presence of misleading narratives in those quotes) necessarily abdicates the responsibility of journalists. Direct quotations still constitute journalistic choice; indeed, in Harry (2014), the author argues that journalistic quotation (direct or

indirect), is a ‘highly interpretive compositional activity’, making it a subjective part of creating a news story. Direct quotations can constitute a *shift* in responsibility to the source as well as a shift in the interpretation or attitude towards the quoted content as not a part of the factual reporting of events (Bergler, 1992, 2006). However, even if this is the case, it is possible that in the context of strategic repurposing of mainstream news, misinformation conveyors, as well as their audience, might be able to ignore this shift.

Future work can also examine the effect of indirect quotation, where quotation markers are not used and instead, quotes are presented using reporting verbs such as ‘said’ (‘they *said* that ... *quoted text* ...’). Such an investigation can help further understand the distribution of where misinformation narratives in mainstream media occur — when citing a source verbatim, or in original writing by the author of the news article.

Our work uses one particular way to categorize outlets as fake news, aiming for precision (see Section 5.2). Future work can also experiment with a broader set of unreliable content (aiming for recall) and identify cases where content of established reliability and unreliability are used together.

We hope that our data, code, and modeling outputs will help enable the investigation of other hypotheses of synergy between certain unreliable and reliable content on social media. This interaction, operationalized by our co-sharing method, can be a valuable tool to help answer different kinds of questions in media studies and sociopolitical information sciences.

By scoring individual news articles for their co-sharing likelihood, our work can also help inspire alternatives to the dominant research paradigm that uses domain-level reliability ratings and tends to treat all individual articles published by the same outlet in the same manner. We note that we also treat articles published

by *fake* news outlets in that manner, but we consider each different individual story published by *mainstream* outlets to identify cases of misinformation adjacency.

Our codebase provides automatically extracted narrative labels and the original raw text data they were extracted from. We encourage user studies and human evaluation to further refine the set of misinformation narratives we introduce in this study in order to make a clearer set of misinformation narratives. This can help answer specific research questions, and conduct additional case studies and investigations of specific misinformation narratives in future research. In Appendix D.7, we provide *all* the high-dimensional narratives automatically extracted from the texts of fact-checked false claims and used in the comparison of their presence in co-shared and control mainstream articles (for the results reported in Table 5.3).³¹

We note that for comparing the presence of each misinformation narrative label in sets of narrative labels derived from the content of co-shared and control articles, this work uses *exact* matching, which, while being reliably precise, can sometimes miss cases of similar narratives that should also be matched such as ‘vaccine injure people’ and ‘vaccine harm people’. While the narrative extraction method clusters terms and phrases using distributional word embeddings, fuzzier matching on obtained labels as an additional step can help ensure that surface-level stylistic differences in the writing of fake news and mainstream news do not hamper the discovery of existing misinformation narratives and false belief propagation in mainstream media. Another smaller future step to control for cases missed by an exact match would be to first lemmatize the terms in all narrative strings or labels, and then conduct our procedure to test our hypothesis.

Finally, we note that contending with other explanations can help flesh out the

³¹Note that these are narratives that are present at least once in one of the co-shared or control group since narrative labels that did not occur in both co-shared and control groups are removed before significance testing (Section 5.4).

relative role of misinformation narratives and also serve as robustness checks for our work. In this work, we already accounted for the effect of partisan curation of news articles (Appendix D.2). However, other possible distinctions between the co-shared set of articles and our control group can be the relative presence of certain entities, or even interest in specific topics that may dominate the fake news world and co-shared mainstream articles. Conducting our hypothesis tests within subsets of articles based on the topic or theme in the news content, and also subsets defined by the presence of prominent entities, constitute interesting directions for future research enabled by our processed dataset.

5.8 Potential Broader Impact

This chapter helps raise general, scientific, and journalistic awareness of looking beyond false claims and statements in fake news content, and critically examining the narratives and larger world views being presented along with the beliefs those falsehoods may support. This helps shed light on how factually correct information might also get repurposed and used to push certain misleading narratives. We direct the reader again to our illustrative example in Fig. 5.1, where the WAPo story content itself makes it clear that since the vast majority of the US population is vaccinated, vaccinated people making up a majority of covid deaths is firmly in line with basic mathematical expectation. However, the original headline (that has since been changed) is ripe for misuse by social media users pushing misleading anti-vaccine narratives (even when the headline is still factually accurate); indeed, our analysis confirms the disproportionate uptake of that article by a number of users known to share fake news regarding vaccines.³² Our case study highlights

³²In addition, we anecdotally observed Twitter sharing of snapshots of just the headline (that was changed in the WAPo article days after the original publication) by a particular popular anti-vaccine

multiple such examples (Section 5.5).

Technology building off of our work can help journalists identify if their content contains existing misinformation narratives and the likelihood their article might get co-opted by conveyors of fake news on social media. Journalists committed to the rigor of their reporting and how their content might get used or misused can adjust their framing accordingly before publishing. This can help reduce the chances of mainstream news articles potentially legitimizing world views based on false or unsubstantiated claims.

We acknowledge that general awareness of our findings could also encourage fake-news conveyors online to more actively seek and share select mainstream articles in support of false claims. However, our study demonstrates that this is already happening and we argue that raising journalistic awareness of this issue in mainstream media is important, as is providing journalists with new tools.

However, there is a potential negative dual-use scenario for the aforementioned new technology — in identifying or scoring mainstream articles for their likelihood of getting co-opted by sharers of fake news, the users who *want* to popularize misinformation narratives could also more readily identify the mainstream articles they can utilize for their goals. In order to mitigate this, it is important that mainstream articles that contain misinformation narratives be reduced, *i.e.*, for the benefit of our technology and findings to be realized quickly through usage in mainstream reporting. We aim to directly convey our findings and recommendations to journalists and facilitate quick uptake of our work and its implications for information sources and platforms committed to the legitimacy of their reporting and content.

account without a link to the full article (and that tweet containing the headline snapshot also underwent disproportionate uptake among the same anti-vaccine fake-news conveyors).

Chapter 6: Conclusion and Future Research Agenda

6.1 Conclusion and Summary of Contributions

Our work engages with text data based on the understanding that language both shapes and is shaped by social reality, society, and culture; therefore, the choices made when communicating in language are socially structured.

Making use of large-scale, often unlabeled, English text data viewed as a digital trace of at least a subset of the social dynamics in the US, this dissertation seeks to address the following question: how can we use unsupervised NLP methods to study texts authored by specific people or institutions in order to effectively explicate the *communicative choices* being made as well as investigate their potential motivations, context-based variation, and consequences?

In Chapter 2, we focused on the method development aspect of the aforementioned guiding question of this dissertation: *explicating communicative choices* in texts using the unsupervised NLP method of topic modeling.

Topic models typically serve as a probabilistic method for the automatic discovery of latent, human-interpretable themes or categories, representing particular communicative choices, from a collection of documents. We first introduced a new topic modeling method that can use the knowledge present in large language models (another powerful unsupervised NLP method) such as BERT to provide a modular addition to any underlying neural topic model and improve the quality of

topic modeling output.

We then conducted a critical audit of how improvement in topic modeling quality is measured or evaluated, and identified crucial gaps in the dominant automated evaluation paradigm used to compare topic models. Through our own experience in topic modeling development followed by a comprehensive survey of the literature, we established that topic models were typically being compared by obtaining one particular automatically computed score (NPMI) for each model. We found that the core assumption behind using these scores — that they were an accurate proxy of human preferences — is invalid, especially in the age of neural topic models that represented a clear shift in the core method architecture from classical models such as LDA.

Our work in topic modeling evaluation connects to critical examinations of automated evaluation conducted for other classes of NLP methods, such as machine translation models, which have also had a tendency to rely on one particular automatically computed score (BLEU) that has been shown to be an inaccurate or invalid proxy of human judgments ([Tan et al., 2015](#); [Smith et al., 2016](#); [Kann et al., 2018](#); [Mathur et al., 2020](#)).

Our findings for topic modeling beckon at reconsidering the evaluation of unsupervised methods and we suggest the following approach:

1. Identify a real-world application or *use case* for the method one wishes to evaluate.
2. Identify the *criteria* that would make the method a good fit for adoption for that particular use case.
3. Develop reliable proxies for measuring methods on those criteria that are grounded in the particular use case.

In [Hoyle et al. \(2022\)](#), we followed the outline above to develop topic modeling evaluation criteria based on the use case of qualitative content analysis, and found that the classical method of LDA fares better on those criteria than the newer neural topic models.

In order to use unsupervised NLP methods to investigate the potential motivations, context-based variation, and consequences of communicative choices, we often linked textual data with information about the authors, social contexts, and media involved in their production in this dissertation.

In our work in Chapter 3, for example, we used knowledge about the speakers or legislators who are the authors of the speeches made on the floor of the US Congress, and linked the texts with non-textual information about the authors such as their organized donors, congressional committee memberships, political party, and home state in the US. While data on both their rhetoric (such as floor speeches) and other non-textual data has been publicly available for a considerable amount of time in our digital age, our linking of these various sources of information over a time range spanning twenty-four years allowed us to examine an understudied connection between money and language use.

Apart from contributing a new, comprehensive dataset, we also shed new light on the role of money in politics. In particular, we showed that financial donations are meaningfully associated with the issues legislators prioritize when speaking in Congress. Our work also revealed the donors connected by policy issues and determined that these relationships and the overall pattern of meaningful associations persisted over time and across congressional sessions. Furthermore, we highlighted how our model can help increase financial transparency by uncovering significantly-connected donation and speech events that occur in close proximity to one another. This work, therefore, provided a concrete example of studying

potential *motivations behind communicative choices* — the choice to give particular amounts of attention and priority to particular policy issues.

In addition to information about authors of texts, our work in Chapter 4 engaged with the particular venues and forums of expression as the non-textual context that can work in tandem with the text itself in order to provide substantial insights into political communication strategies of elected representatives in the US. We used an existing unsupervised NLP method that quantifies the ideological positioning of a set of authors using their *communicative choices* in the form of ideological framing and usage of polarized terms within political issues.

After extensive validation of this existing tool and tweaks to enhance modeling stability and flexibility, we contributed a new understanding of *variation* in ideological positioning based on heterogeneous audience-based considerations (or the *context*) across voting in Congress, speaking in Congress, and communicating on social media (Twitter). We also constructed and contributed an original dataset combining various sources of information about members of Congress with three different ideal point measurements, derived from their voting records on bills in Congress, their floor speeches, and their tweets.

Our works in Chapter 3 and Chapter 4 included human (expert) curation of the outputs of a topic model. Domain experts used the top words and top documents to provide labels for automatically discovered topics. In some cases in Chapter 4, they also provided various kinds of annotations for the discovered concept such as coherence ratings and expectation that a particular policy issue will be discussed in a polarized manner in the US Congress. We release software that can use the outputs of a topic model — such as the processed vocabulary and the topic-word distribution — and create spreadsheets to facilitate curating labels and other annotations for discovered topics at https://github.com/Pranav-Goel/Topic_Model_Curation.

Finally, recognizing that including particular narratives of the form of entity-action relationships such as ‘vaccinated people spread virus’ or ‘illegal immigrant commit crime’ represents *communicative choices* made by journalists in news articles, we contributed new understanding to mainstream media’s role in online misinformation (Chapter 5). To enable this particular research and contribution to social science and digital journalism, we made use of publication outlet metadata and social network data as the non-textual contextual components in order to obtain mainstream news articles that are co-shared with fake news content (and also obtain a corresponding mainstream news control group). We found that a significantly higher presence of misinformation narratives is one factor in the disproportionate uptake of mainstream news articles by conveyors of fake news on Twitter; this can be viewed as a *consequence* of choices made in the particular subset of mainstream articles.

Our findings also suggested a strategic repurposing of mainstream news by conveyors of misinformation as a way to enhance the reach and persuasiveness of misleading narratives. Our work has important implications for responsible journalistic practice and also represents a significant paradigmatic difference from most current scholarly work on misinformation, which focuses on false news reports. The modeling outputs of our work create new categorizations of existing mainstream news articles into those that are co-shared with fake news and those that are not — this new data will enable further research into mainstream media’s potential role in online misinformation.

6.2 Future Research Directions

In addition to specific steps and ideas for future work highlighted in each of the Chapters 2 to 5, we now discuss future research directions that both tie into the general theme of this dissertation as well as build off of our particular research works and contributions.

6.2.1 Systematic *Exclusion* of Certain Communicative Choices

In all the prior chapters, we used texts authored by various actors who are part of a particular ecosystem or institution — such as elected governmental representatives or news media sources — to study the communicative choices being made by them, along with their potential motivations, variation, and downstream consequences. To study these choices, we followed what can be considered a common paradigm in unsupervised NLP research and identified all the choices being made in a particular collection of documents.

This form of operationalization for the detection of choices made by actors in textual data implicitly considers the entire set of possible choices as ones that can be extracted from all the texts in a given collection, authored by a particular set of actors often belonging to the same institution or ecosystem. However, there might be other possible choices that will not be considered in this paradigm because they are not made by anyone in that ecosystem. In other words, certain communicative choices might be systematically excluded or absent when texts authored by only the actors belonging to a particular institution are examined. When those institutions hold great influence — as elite institutions like politicians and news media do in terms of agenda setting and framing of discussions (Iyengar and Simon, 1993; Entman, 1993; McCombs, 2002) — it is important to find and analyze the possibilities *not* captured

in choices made by those actors.

Centering this understanding of possible exclusion or absence of possible choices in one collection of texts, in future research, we plan to develop a method that can use two different collections of texts to understand possible communicative choices made in one collection but not the other. Our sociopolitical motivation would be to highlight missing choices in the elite discourse on an issue, *i.e.*, the choices made by *none* of the actors belonging to a particular elite institution in a given context, when those choices *were* made in corresponding public discourse. This can help future work in social sciences hold institutions accountable if and when certain options, perspectives, concepts, or ideas are not even being considered in the discussions and discourse they author, when they might be relevant.¹

Concretely, this future research agenda is motivated by prior empirical research comparing public and elite discourse in the social sciences, especially research that utilized text data. For example, [Kortenska et al. \(2020\)](#) used newspaper statements (to operationalize elite discourse) and prior research on salient categories and frames known to be present in the public citizen discourse around European integration in Serbia, and conducted manual content analysis to create codes or concepts (represented using sets of keywords). They then compared the two sets of concepts to find differences in discourses among the citizens and elites in Serbia (referred to as the ‘public-elite gap on European integration’). [Sunderland et al. \(2009\)](#) compared the elite and public discussions around disability, with the former operationalized using ‘small samples of government and supranational policy documents’ and the

¹Here we use the term *discourse* based on the definition provided in [Sunderland et al. \(2009\)](#): “Discourse ... is any form of communication or social interaction that produces and reproduces shared meanings and understandings.”; and as [Sunderland et al.](#) note, *discourse* does not have to be limited to text or any one form of communication, though in our case, text will likely serve as the formal or operational form.

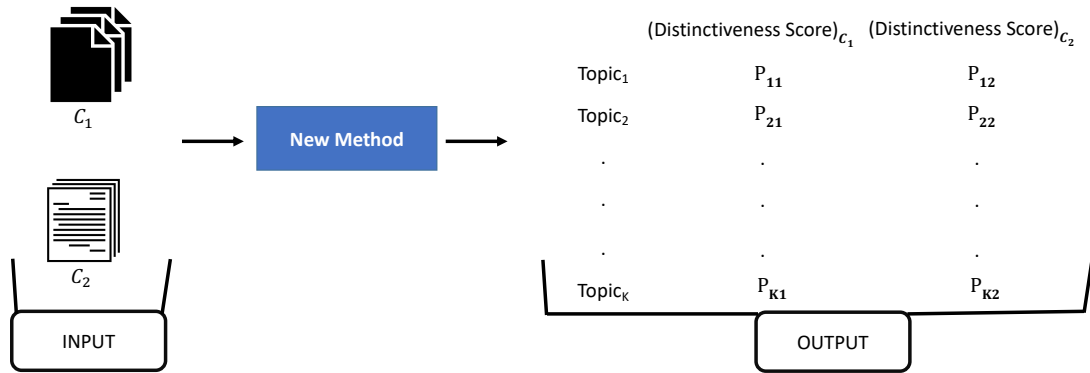


Figure 6.1: The objective of the new topic modeling-based method we plan to propose in future research is to help humans (who are studying a particular discourse) find distinctive topics that are present in one but not the other collection; where authors of the two collections are asymmetrically related. C_1 and C_2 represent the two text collections.

latter using ‘life story interview transcripts and focus group transcripts’ from people with acquired disability. They hypothesized that discussions or discourse around joy and happiness are *missing* in elite discussions around disability. Validation of this hypothesis was conducted using word lists of happiness and joy-related terms, and by quantifying their presence in elite versus public collections of texts.

The most directly connected prior work that can also help provide a concrete dataset for this future direction is [Chen et al. \(2022\)](#), which studied the differences in public and elite climate change discourse, specifically around climate change strikes and organized movements. The public discourse was operationalized using Twitter tweets from climate change activists and organizers (and how they portray their efforts), and the elite discourse was operationalized using news media articles on climate change movements. Using two separate topic models to discover concepts in the two collections, [Chen et al.](#) found key concepts in the public discourse missing from climate change action coverage in mainstream news media such as social and environmental justice, highlighting of collective action victories, and emphasis on protests.

From a dataset perspective, our future work will distill the aforementioned broad motivation to one key component as a minimum criterion for the data: *asymmetry* between the groups of actors who author the two different collections being compared. Asymmetry is arguably emblematic of the relationship between members of public and actors in elite institutional spaces such as news media and elected politicians. Such a setting would represent a contrast to much of the prior work in this research area that investigates differences in collections of texts authored by groups largely operating within the same institution and are symmetrical in their relations, such as Democrats versus Republicans, liberal vs conservative news outlets, or fake news outlets vs mainstream media (such as our own investigation in Chapter 5). This work would aim to construct new datasets consisting of collections of texts that instantiate relevant elite and public discourse, and also utilize existing datasets where the authors present in the two different collections of texts share an asymmetrical relationship.

From a methods perspective, this work will adapt existing techniques as well as contribute new methods that can compare *two* collections, with the understanding that one straightforward way that such a contribution can be applied to more than two collections is by conducting multiple pairwise comparisons. To further define the intended methodological scope of this work, we will seek to analyze ***latent concepts as the communicative choices***. This would contrast with prior textual analysis literature that compares collections in the *observable* space, often using information-theoretic approaches (such as divergence measures or goodness-of-fit statistical testing) in order to find words or phrases that are characteristic of one collection when compared with another collection. [Monroe et al. \(2008\)](#) provide a classic example of this ‘differences in the observable’ research in political science analysis literature, looking at partisan terms by comparing texts authored by

Democrats and by Republicans in the US. [Demszky et al. \(2019\)](#) similarly contribute to understanding polarization by comparing the choice of topics, framing, affect, and illocutionary force in tweets on mass shootings by Democrats and Republicans. For a more recent generalized approach that enables a suite of different methods as well as visualization for pairwise comparison of texts in the observable space, see [Gallagher et al. \(2021\)](#).

With the goal being the unsupervised discovery of latent concepts from a collection of texts, **topic models** present themselves as a natural choice of methodological focus for this future work, and this also allows us to build on our extensive prior experience in this methodological space (Chapter 2).²

Fig. 6.1 provides a straightforward visualization of the intended input and output for our potential new topic modeling-based method. The main objective of the proposed new methods in this future research can be stated as **helping humans find meaningful concepts that are present in one collection of texts but not the other, where the two collections are authored by groups with an asymmetrical relationship such as the public and the elite**. These concepts can be referred to as *distinctive concepts* or topics.

In newer scenarios represented by data not used in this future research itself, these humans or end-users will often be domain experts and research scientists who can use the findings of our bottom-up approach to validate hypotheses in a particular context. For datasets that get utilized in this research, expert or crowd-sourced human annotations will help serve as the ‘ground truth’ for training our methods, as well as for the evaluation of the automated methods.

²Apart from topic modeling, advances in large language models may also present a viable path for our goals. For example, using an iteration of the GPT-3 model ([Brown et al., 2020](#)) that can utilize human feedback ([Ouyang et al., 2022](#)), [Zhong et al. \(2022\)](#) introduce an approach to elicit natural language hypotheses to explain differences in the distributions of a pair of text collections.

Specifically, we intend for this future research to contribute generalizable and effective methods for automated and human-in-the-loop discovery of distinctive concepts, which can then be used to both validate known differences in two collections authored by asymmetrically related groups (such as public and elite), and also be used to find potential gaps that merit further investigation by relevant experts.

6.2.2 Beyond Individual Choices: Using Similarities in Communicative Choices to Model Agreement in Decision-Making for Pairs of Actors

Motivation. It is common in computational social science to model individuals' decisions, such as a legislator's vote on a bill, or the choice a user makes online to share or not share a social media post (Nguyen et al., 2015b; Kraft et al., 2016; Kornilova et al., 2018; Barberá, 2015; Hong et al., 2013; Metaxas et al., 2015; Meier and Elswailer, 2019). Computational modeling of votes on congressional bills has helped inform ideal point models (Poole and Rosenthal, 2000), and going beyond just vote-centric data, researchers have developed models that include unstructured language data in order to provide a more nuanced picture (Gerrish and Blei, 2011; Gu et al., 2014; Lauderdale and Clark, 2014; Sim et al., 2015; Nguyen et al., 2015b; Vafa et al., 2020) as we noted and ourselves utilized in Chapter 4. A more generalized view of decisions as votes has yielded insights into public attitudes by, for example, modeling account-following choices on Twitter (Barberá, 2015; Gu et al., 2016; Pauls et al., 2015; Demszky et al., 2019) and looking at discrete real-time audience responses to political debates (Argyle et al., 2019).

Although the majority of such models focus on choices at the individual level, some work has attended to the fact that many if not most decisions of this kind

take place in the context of a *population* of individuals. In such settings, additional information about relationships between individuals may yield additional predictive power and insight. For example, traditional theories of homophily suggest that shared properties like party membership, legislative committee assignments, electoral geography, or gender will increase the likelihood that two legislators will make the same choices ([McPherson et al., 2001](#); [Kirkland, 2012](#); [Clark and Caro, 2013](#); [Baller, 2017](#); [Wojcik, 2018](#); [Fischer et al., 2019](#)), and, conversely, looking at whether and when people vote together can yield insight into the factors that influence decision-making. This reframing of the core behavior of interest — from “how does an individual vote, and why?” to “do these individuals vote the same way, and why?” — defines an emergent shift of research questions from *voting* to *co-voting* ([Ringe et al., 2013](#); [Peng et al., 2016](#); [Wojcik, 2018](#)).³

Goals. Primarily, we will seek to extend the co-voting framework introduced by [Ringe et al. \(2013\)](#) to incorporate analysis and comparison of individuals’ language into the model. Conceptually, the comparison of two individuals’ language, in terms of the issues they discuss and the framing of those issues, will be treated as a pairwise relationship playing a role in homophily that is directly analogous to the role of other factors such as representing the same state or being members of the same Congressional caucus. Our preliminary results below show that sophisticated modeling of language (such as the TBIP method we discussed and utilized in our research in Chapter 4) and language-use-based similarity of a pair of legislators holds the potential to model co-vote agreement better than just information about the party or home state.

³Our motivation behind exploring the co-voting or co-decisions framework along with the use of this framework proposed in this section have already been adapted for a substantive application for a new NLP method in [Hoyle et al. \(2023\)](#).

We will also look to address more general use cases to validate the utility of language or communicative choices beyond the political domain, moving from modeling of co-voting to modeling *co-decisions*, where a decision is a generalized vote. Consider, for example, applying co-voting models in the context of citations in the scientific literature, where $\Pr(\text{co-decision}(a_i, a_j | r, w_i, w_j, \mathbf{X}_{i,j}; \theta))$ is the probability that authors a_i and a_j , writing paragraphs w_i and w_j , respectively, will both cite reference r , in a model parameterized by θ that takes into account pairwise features $\mathbf{X}_{i,j}$ (such as frequency of previous co-authorship or membership in the same professional organizations). Many use cases, including this one, are isomorphic to the legislative co-voting case: here the a_* correspond to the legislators, w_* to their speeches about a bill r , and $\mathbf{X}_{i,j}$ captures pairwise features such as membership in the same party or caucus. Utility in other use cases will gain leverage from a key insight, which is that the *prediction of a co-vote* for two individuals can also be interpreted as a *recommendation*, if one of the individuals' votes is already known. For example, if the co-decision probability above is high, and we know that author a_i cited r , it makes sense to *recommend* to author a_j that they cite r in this context also. We discuss how our prior work on misinformation and news-sharing behavior on Twitter in Chapter 5 can be viewed in this co-decisions framework as a concrete motivation for this general future research direction below.

To this end, using co-vote as the initial application, we have packaged our implementation of a co-decisions framework and provided it as publicly available software, in order to spur both our own planned future research and other future work carried out by social scientists.⁴

Finally, we will look to use this framework in order to propose new innovations in NLP and modeling language in general, in particular, better methods to capture

⁴<https://github.com/Pranav-Goel/co-decisions>

similarity in language use and communicative choices of pairs of actors. These advances in language modeling will be validated formally by incorporating them into pairwise comparisons of legislator language and demonstrating improvement in performance within the generalized co-voting framework above as part of our future research.⁵

Initial Formulation. Following [Ringe et al. \(2013\)](#) and [Wojcik \(2018\)](#), we will begin by modeling the log odds ratio of the *co-voting rate* between a pair of legislators i, j using a mixed effects regression, which controls for the random effects of both actors under consideration. The co-vote rate is the number of times the legislators vote the same way — yea or nay — divided by the total votes over a period.

$$\mathbb{E} \left[\log \left(\frac{(\text{Co-Vote Rate})_{ij}}{1 - (\text{Co-Vote Rate})_{ij}} \middle| \mathbf{X}_{ij}, a_i, b_j \right) \right] = \beta_0 + \boldsymbol{\beta}_k^\top \mathbf{X}_{ij} + a_i + b_j \quad (6.1)$$

$$a_i, b_i \sim MVN(0, \Sigma); \quad \Sigma = \begin{bmatrix} \sigma_a^2 & \rho \sigma_a \sigma_b \\ \rho \sigma_a \sigma_b & \sigma_b^2 \end{bmatrix}$$

where a_i and b_j model random effects for legislator i and j .

Random effects help deal with the fundamental problem of social network correlation ([Holland and Leinhardt, 1981](#); [Anagnostopoulos et al., 2008](#)), since pairs in a network cannot be considered independent.⁶

$\mathbf{X}_{ij}$ is an n -dimensional vector of features, where each feature captures a type of pairwise relationship between legislators i and j and serves as the entry point for much of our innovation, both in terms of incorporating language and in terms of generalizing beyond Congress. These can include binary features for state

⁵Prior NLP research has proposed models that directly model *pairs* of texts in order to measure specific things about the pair such as the semantic similarity; for example, see [Amiri et al. \(2016\)](#).

⁶Following [Ringe et al. \(2013\)](#), we will discard cases of complete agreement or disagreement.

membership, party affiliation, and committee membership; or continuous features for the similarity between their constituents' ideology (Miller, 2016), real-life social network statistics (Ringe et al., 2013), Twitter connections (Wojcik, 2018), bill co-sponsorship ties (Kirkland, 2011; Fischer et al., 2019; Gross and Kirkland, 2019), or joint press releases (Desmarais et al., 2015).

We will leverage decision makers' (in the co-vote case, legislators') language in this co-vote or co-decision framework. The pairwise matrix X_{ij} has generally been used in previous work to model social relationships between legislators i and j . However, our key insight is that it provides a very general mechanism for incorporating *any* set of features describing a pair of legislators: not just structured knowledge, but information derived from *unstructured* data such as language. The general framework is to use some measure of similarity between the language of legislators i and j .

Variable	Estimate (Std. Error)
(Fixed Effects)	
Intercept	0.2906 (0.0238)
Same Party	0.8608 (0.0018)
Same State	0.0554 (0.00366)
Similarity between Speech Vectors	1.1192 (0.0348)
(Random Effects Variance)	
Legislator 1	0.0786
Legislator 2	0.1007

Table 6.1: Preliminary results for our basic mixed-effects model (Eq. (6.1)) using just three pairwise features or fixed effects (X_{ij}): same party, same state, and the cosine similarity between the TF-IDF vectors formed from the text of floor speeches for the pair of legislators. The dependent variable is the co-vote agreement rate for the 109th Congress. The estimates indicate that even using very basic speech-based features can offer a meaningful association and explanation for the co-vote rate between two legislators. As expected, the more similar the speech vectors, the higher the rate of voting together.

Preliminary Results and Contribution. As a concrete starting point, we used the cosine similarity of the TF-IDF vector representations of the legislators’ congressional floor speeches and find promising initial results (Table 6.1). This utilizes speech and voting data from the 109th Congress.

Variable	Estimate (Std. Error)
(Fixed Effects)	
Intercept	-1.0795 (0.04)
Same Party	0.8288 (0.0019)
Same State	0.056 (0.0036)
Similarity between Speech Vectors (TF-IDF)	0.8852 (0.0351)
Similarity between Legislators’ Speech-based Ideal Points	1.5783 (0.0368)
(Random Effects Variance)	
Legislator 1	0.0774
Legislator 2	0.0945

Table 6.2: Additional preliminary results for our mixed-effects model (Eq. (6.1)) using just four pairwise features or fixed effects (X_{ij}): same party, same state, the cosine similarity between the TF-IDF vectors formed from the floor speeches for the pair of legislators, and similarity between the estimated ideal point values for legislators derived from their floor speeches using the TBIP method. The dependent variable is the co-vote agreement rate for the 109th Congress. The estimates indicate that the closeness of speech-based ideal point values for legislators has a significant impact on the co-vote agreement rate, more than sharing the same party or state and even more than our basic modeling of speech-based legislator similarity using TF-IDF vectors. This further shows the promise of using language-based similarity between decision-makers, and how more sophisticated language-based tools can add further value to co-decision modeling.

Further, using the TBIP tool (Vafa et al., 2020) to obtain ideal point values using just the speeches given by legislators (as explained and utilized to study context-based variation in communicative choices in Chapter 4), we find that using the distance between these text-based ideal point values offers an even higher estimate for modeling co-votes (Table 6.2).

These initial results showcase the potential of using similarity in language use and communicative choices for better modeling of agreement or disagreement in decision-making for pairs of actors.

Our implementation of the above formulation and example application to co-voting is publicly available at <https://github.com/Pranav-Goel/co-decisions>.

Understanding our Work on Capturing Misinformation-Adjacent Mainstream Articles as an Application of the Co-Decisions Framework. Sharing any URL (link to an external website) on social media sites (such as Twitter) can be seen as a *decision*. Millions and billions of such decisions are being made on Twitter every day. Our co-decision framework can help answer substantive questions on the online information ecosystem as seen through the prism of sharing behavior on social media. Since the decision in this context is sharing, our co-decision framework maps onto *co-sharing*; a concept we utilized to study consequences of communicative choices in Chapter 5.

Specifically, recall that in Chapter 5, we examined misinformation narratives; misleading narratives that pull together content from a myriad of sources to support incorrect conclusions. We evaluated the presence of misinformation narratives by looking at *co-sharing* of content from fake news sites with content from mainstream news as a way to detect the alignment of select mainstream news with fake news. This can be viewed as a concrete application of *our generalizable co-decision framework* in order to find relevant content from mainstream news.

We found evidence that articles from reliable outlets *co-shared* with misinformation contain a significantly higher presence of misinformation narratives. Our findings suggested a strategic repurposing of mainstream news by conveyors of fake news as a way to enhance the persuasiveness of misleading narratives.

6.2.3 Studying Variation in Ideological Expression across Political Issues

In this future research direction, we aim to explore an extension of our work studying context-based variation in ideological expression and positioning in Chapter 4. There, the context was the venue and audience-based considerations. Another critical context for ideological positioning is the particular policy issue under discussion, and we will aim to develop a robust, validated NLP tool to derive issue-specific text-based ideal point estimates.

A limitation of the conventional approach to measuring legislative ideal points is that they are measured in the aggregate and ultimately denote a single scalar value for each legislator — this is true for traditional vote-based ideal points (Poole and Rosenthal, 1985, 1991), and also for the ideal point measurements derived from text using TBIP (Vafa et al., 2020) as explained in Section 4.5. While such a measurement has the benefit of greater simplicity in providing a singular summary of a legislator’s ideological position, it masks important variation across issues (Bateman and Lapinski, 2016).

Focusing on smaller issues in appropriations bills in the US House, Crespin and Rohde (2010) find multidimensional ideological positioning for legislators. Empirically, Gerrish and Blei (2012) develop an issue-adjusted vote-based ideal point model, and Moser et al. (2021) also offer such an extension in order to discover different ideological preferences in different policy domains; both these works do find variation in vote-based ideological positioning of US legislators. We expect similar variations across issues or policy areas can be exhibited across issues in political rhetoric and communicative choices; indeed, Nguyen et al. (2015b) developed a joint model of votes, bill texts, and floor speeches and, using a hierarchical

topic modeling-based method, highlight differences in framing and support for a particular faction within the Republican party across various policy issues.

On an individual level, legislators have variable levels of personal interest and preferences regarding different issues ([Van Houweling, 2003](#)). In addition, at a group level, scholars argue that some issues are “owned” by one party or the other, which indicates that there are issues on which a party is advantaged in the eyes of voters; a classic example of issue ownership includes Democrats owning education policy and Republicans owning defense issues ([Petrocik, 1996](#)). [Egan \(2013\)](#) find that parties are ideologically more rigid on issues that they own. The intensity of inter-party disagreement can vary across different policies or issues — [Cayton \(2021\)](#) find that the parties disagree more strongly on the issue of taxation, but not as much on the issue of regulation and communication in terms of stances on the functions of a government. [Cayton](#) also emphasizes that while the singular economic axis captured by traditional vote-based estimates such as DW-NOMINATE is indeed the most salient in their analysis, smaller issue-specific dimensions also exist. Further, the ideological positioning of parties can vary in different ways across different issues like abortion and civil rights in response to pressure from interest groups and activists ([Bawn et al., 2012](#)).

Factions within a party also seek to organize and differentiate policy preferences from the rest of their party ([DiSalvo, 2009](#); [Nguyen et al., 2015b](#)) — while such groups tend to publicly advertise their policy positions and differences from their party, the self-expressed differences in policy agendas might differ from empirical findings in the data including textual data, or might change over time. Further, some groups and caucuses based on demographic identity might exhibit ideological positions on some issues that are different from the ‘mainstream’ party position.

For political actors other than legislators, some recent content analysis and NLP

studies have found multidimensional ideological spaces; for example, news media outlets in [Sinno et al. \(2022\)](#) and social media users in [Monti et al. \(2021\)](#). Ideological positioning and belief systems of the mass public have also been determined as multidimensional, with many US citizens holding different liberal or conservative preferences across different policies and issues ([Treier and Hillygus, 2009](#)). Examining the ideology of the American electorate with respect to domestic politics, multiple prior works posit *at least* a two-dimensional view (economic and social) of ideological positioning ([Carsey and Layman, 2006](#); [Carmines et al., 2012](#); [Feldman and Johnston, 2014](#)). Given our own results in Section 4.6 showing possible responsiveness to constituent preferences, there is good reason to expect that legislators’ ideological positions may differ across issues and that this will be reflected in their floor speeches and tweets since these communiques are public-facing in nature.

We will use the above-elucidated motivation to introduce a new model to derive issue-specific estimates for authors from text.

In a preliminary attempt, we tried to introduce a simple extension to the existing TBIP method we successfully used to contribute a new understanding of legislative communication dynamics in Chapter 4. Recall the TBIP formulation shown in Section 4.5 and Eq. (4.2):

$$y_{dv} \sim f(\sum_k \theta_{dk} \exp(\log \beta_{kv} + x_{a_d} \eta_{kv})) \quad (4.2)$$

In our attempt, instead of a scalar ideal point x_{a_d} for the author a of a document d , we made a small change to the method in order to allow it to infer a k -dimensional vector of ideal point values to represent each author a using one ideal point value for each topic ($x_{a_d k}$). This transforms the original formulation above as:

$$y_{dv} \sim f \left(\sum_k \theta_{dk} \exp (\log \beta_{kv} + \mathbf{x}_{\mathbf{d}\mathbf{k}} \eta_{kv}) \right) \quad (6.2)$$

In theory, this can allow the inference of a multidimensional ideal point space where one ideal point value per topic is inferred for each author, instead of just one aggregated scalar. To ground this model in its practical implementation, we used non-issue-specific TBIP estimates as a prior on the distribution used to model issue-specific estimates.

However, while an initial small synthetic data-based proof-of-concept was able to demonstrate that the theoretical aim *can* be achieved by our extension, a comprehensive human validation effort on the actual real-world dataset of floor speeches across the 115th and 116th Congresses revealed that the model fails in important ways and is not yet viable for practical use. Specifically, we reconducted our human annotation experiments from Chapter 4 (results provided in Appendix C.3) and conducted a head-to-head comparison between the original, non-issue-specific TBIP model and our issue-specific extension. We found that there is a significant reduction in the number of topics where the ideological framings at the two ends of the spectrum are interpreted as an aspect of the original, neutral issue or concept. There is also a significant reduction in the number of topics for which two coherent, ideologically polarized framings are discovered per expert judgment.

Therefore, informed by our experience in validating one straightforward extension to the tool, we will look to take a different approach for modeling issue-specific ideal point estimates in future work. Advances in large language modeling such as the introduction of GPT-3 (Brown et al., 2020) can offer one potential pathway for unsupervised discovery of rankings across legislators on a specific issue, since these models can capture a construct based on a few samples of text.

One issue for our above-failed attempt at extending TBIP is that the model is severely underspecified when aiming to discover ideal point estimates across all legislators and across all discovered issues. Furthermore, while there is value in unsupervised modeling, in the case of legislators and congressional data, voting records alongside speeches or texts *are* available and can help guide ideological positioning estimation as was the case in [Nguyen et al. \(2015b\)](#). In future work, we will also seek to build new models that jointly model vote as well as text in order to provide a more robust ideal point estimation, combining the standard, well-studied nature of voting in Congress with the nuance offered by language use.

6.2.4 Estimating Causal Association Between Donation Activity and Legislators' Communicative Choices

With the new, comprehensive dataset combining data on congressional floor speeches, donations data, and information about legislators across twenty-four years that we introduced in our work in Chapter 3, we aim to use existing tools and potentially develop new ones in order to examine the question of causality. This attempt will help estimate the relative occurrence of donations influencing legislators' issue-attention versus legislators' communicative choices attracting newfound donor support. As mentioned in Section 3.8, in this future research, we will focus on legislators who 'break character' by talking about an issue they generally do not pay attention to in a particular period of time (within a particular congressional session), and systematically analyze if donors interested in that issue tend to donate to the legislator before or after those unexpected speeches.

Concretely, our dataset contains dates where speeches are made as well as dates for each transaction tracking contributions from donors. We also obtained annota-

tions from domain experts using automated modeling outputs. This combination of modeling outputs and expert annotations can help us divide donors by relevant policy issues where a strong association between speeches on a particular issue and that donor's activity has already been established. By focusing on one particular Congress at a time, we can control for at least some extrinsic factors. And instead of necessarily trying to obtain *one* aggregated answer for the age-old causality question in the realm of donations and congressional activity (instantiated via language use), we will instead allow for different possibilities for each legislator-donor pair, *i.e.*, we will seek to establish how well cases of legislator demonstrating sharp changes in issue-attention is either preceded by or followed by cases of a donor changing their usual pattern of donation to that legislator.

A major challenge with attempts to establish causality is the possibility of both variables responding to some extrinsic factor simply not present in the data. We acknowledge this concern applies to the aforementioned approach for our future work. We will seek to adapt methods to discover cases where a big, timely donation from a donor influence the speech patterns of the recipient legislator in ways that would *not* have happened without that donation.

6.2.5 Investigating Possible Contrasting Choices in Opinion Pieces (Op-Eds) and News Published by the Same Popular Mainstream Media Outlets

In Fig. 5.2 in Chapter 5, we highlighted the top ten articles from both the New York Times (NYT) and the Washington Post (WAPo) that have the highest score for being co-shared with fake news on Twitter in our dataset. We only showcased the news items and removed opinion pieces, but for both outlets and especially

WAPo, the top articles co-shared with fake news include opinion pieces or *op-eds*. The top three such co-shared op-eds published by WAPo are: “Opinion: Explain the Chinese spy, Sen. Feinstein”, “Opinion: No, the White House didn’t ‘dissolve’ its pandemic response office. I was there.”, and “Opinion: Someone committed a crime in the Michael Flynn case. It wasn’t him.”.

Op-eds have been shown to be influential in terms of changing the mind of the general public and policymakers across the ideological spectrum in a long-lasting manner (Coppock et al., 2018). Scientists and academics, for example, have considered op-eds a valuable tool to disseminate their research to the public as well as to influence policy (Sommer and Maycroft, 2008; Besley and Nisbet, 2013).

We will aim to take a closer look at op-eds in the dataset we created containing articles from mainstream news outlets that are co-shared with fake news as well as those articles that are not co-shared (Chapter 5) and examine whether op-eds, especially for popular outlets such as NYT and WAPo, are present at a significantly higher rate in the co-shared set of articles. We will use this analysis to then examine the possible role of mainstream platforms’ op-eds in online misinformation.

Apart from a higher presence of misinformation narratives (Section 5.4), another complementary and significant factor that characterizes mainstream articles co-shared with fake news is their conservative bent, in terms of relatively higher shares by Republicans on Twitter (Appendix D.2). A particular mainstream outlet can and does publish an ideologically heterogeneous set of articles (Green et al., 2021). The following reporting and analysis from 2017 — <https://www.cjr.org/criticism/bret-stephens-op-ed-new-york-times-wall-street-journal.php> — suggests that conventional thinking supporting status-quo systems and conditions such as capitalism and inequality (in terms of wealth, and power more broadly) often finds its biggest defenses in op-eds published by mainstream outlets generally perceived

as liberal-leaning (in the aggregate).

Motivated by the literature and our own anecdotal observations, we wish to examine the ideological bent of op-eds in this future work. We will consider one mainstream outlet at a time such as NYT and WAPO, and compare the ideological bent of op-eds published by that outlet over the same time period. This comparison can also be conducted by considering the politics expressed in just the op-eds versus the general, aggregate ideological bent of that news outlet.

We plan to include at least three measures of ideological leaning in such an analysis: *a*) Twitter shares on social media (Republican versus Democrat shares; following [Green et al. \(2021\)](#) and the proxy utilized in our work in Chapter 5); *b*) text-based measures using the content itself for deriving ideological or partisanship scores, building off of methods such as TBIP ([Vafa et al., 2020](#)) as utilized in our work in Chapter 4; and *c*) expert annotation with political scientists (whose own political ideology will also be noted following [Yano et al. \(2010\)](#)) who would be asked to rate samples of op-eds as well as news story texts (without being provided these labels explicitly) — these expert ratings can then be aggregated to compare news articles and op-eds.

Apart from conservative versus liberal quantification of op-eds, we will also look to analyze the rate at which op-eds defend the status quo (regardless of ideology) versus advocate for change, especially radical change. Our hypothesis is that in a given time period, op-eds from an outlet are more conservative and defensive of the status quo than the average news reporting.

We will also seek to study the *timing* of op-eds in relation to the mass discourse and potential protests and policy calls pushing for changes — assuming we find a systematic conservative bent in op-eds, are they strategically published at the same time as a popular or influential push for change in the US? As specific case

studies, we will examine op-eds from NYT and WAPO in the summer of 2020 around the Black Lives Matter protests as well as op-eds around the time of mainstream media coverage of wealth tax and general progressive taxation policies pushed by Sen. Elizabeth Warren and Sen. Bernie Sanders especially in late 2019 during the Democratic party primary debates. With this, we will study the *possible* role of opinion pieces platformed by major mainstream outlets in reinforcing the status quo, especially during the time the rest of the news reporting at that same outlet might be covering an agitation for progressive changes.

6.2.6 Does Misinformation Benefit from a Partnership with Expression of Targeted Hatred?

Using the base provided by our work in Chapter 5 including Twitter sharing behavior around various fake news articles, extracted misinformation narratives, and tweet content sharing fake news, we will try to understand the potential role of targeted hatred and vilification in online misinformation spread. A false claim or a fake news article can have a particular policy, idea, or concept at the core of the misinformation, or center an individual or group of humans. Our investigation will aim to understand if targeting people (as particular entities) for hatred, vilification, or dehumanization leads to a higher likelihood of misinformation spread online.

One reason such an investigation is interesting is the potentially contrasting benefits of the decision to include or exclude negative sentiment or hatred towards a particular entity. On the one hand, targeting particular entities or dehumanizing the out-group can help solidify in-group support and lead to a higher chance of in-group adoption of the misinformation narrative. The in-group might accept or be apathetic towards the lack of accuracy in order to bolster their out-group

animus and beliefs. Disinformation has been shown to often carry an agenda of increasing opposition and ill-sentiment towards out-groups and minorities ([Bennett and Livingston, 2018](#); [Marwick and Lewis, 2017](#)). On the other hand, perhaps targeting a particular concept or idea without including hatred or targeting any particular group including those who disagree about that (false) claim could help the claim spread such that it transcends community boundaries, leading to more widespread adoption or at least a lack of effort to challenge it from any group, inviting a passive but wider persistence of the claim.

Previous scholarly work has examined the relationship between disinformation campaigns and hate speech ([Reppell and Shein, 2019](#)), how targeted misinformation can function as an aspect of hate speech in order to bolster far-right narratives that hurt minorities and marginalized communities ([Darmstadt et al., 2019](#)). Targeted hatred against Asian people and communities due to misinformed narratives during the early part of the ongoing COVID-19 pandemic in 2020 constitutes an iteration of such a relationship between misinformation and hate speech ([Kim and Kesari, 2021](#)). [Humprecht \(2019\)](#) found that “... in Germany and Austria, disinformation frequently targets immigrants.”

Recent work in [Hameleers et al. \(2022\)](#) serves as the most concrete motivation for this future work, where the authors examine hate speech in fact-checked false claims and “... postulate that the interaction between communicative untruthfulness and hate speech corresponds to the politics of disinformation.” [Hameleers et al.](#) find that “hate speech, partisan attacks, and other forms of incivility” can serve as characteristic traits of disinformation content.

Our work and data in Chapter 5 provides new tools for examining the relationship between misinformation and hate speech or dehumanization, especially as a tactic that can influence misinformation *spread*.

Appendix A: Extracting Communicative Choices from Textual Content: Enhancing and Comparing Topic Models

A.1 Dev Set Results

We optimized our models on the dev set, froze the optimal models, and showed the results on the test set in Tables 2.2 and 2.3.¹ We show the corresponding dev set results for those models in Tables A.1 and A.2.

	20NG	$K = 50$ Wiki	IMDb	20NG	$K = 200$ Wiki	IMDb
DVAE	0.341	0.512	0.137	0.312	0.470	0.155
W-LDA	0.294	0.500	0.136	0.203	0.310	0.095
SCHOLAR	0.343 (0.003)	0.504 (0.007)	0.167 (0.002)	0.279 (0.002)	0.478 (0.005)	0.139 (0.002)
SCH. + BAT	0.377 (0.006)	0.526 (0.009)	0.180 (0.002)	0.343 (0.002)	0.518 (0.001)	0.174 (0.003)

Table A.1: The development-set NPMI for our baselines (Section 2.1.3) compared with BAT (explained in Section 2.1.1) using SCHOLAR as our base neural architecture. We achieve better NPMI than all baselines across three datasets and $K = 50, K = 200$ topics. We use 5 random restarts report the standard deviation.

	20NG	Wiki	IMDb
W-LDA	0.294 (0.014)	0.500 (0.013)	0.136 (0.009)
+ BAT	0.316 (0.010)	0.511 (0.016)	0.162 (0.003)

Table A.2: The mean development-set NPMI (std. dev.) across 5 runs for W-LDA and W-LDA + BAT for $K = 50$, showing improvement on all datasets. This demonstrates that our innovation is *modular* and can be used with base neural topic models that vary in architecture.

¹The entire contents of Appendix A.1 are directly taken from Hoyle et al. (2020).

A.2 Extrinsic Classification Results

The primary goal of our method is to improve the coherence of generated topics.² It is natural, however, to ask about the impact of our method on downstream applications. We include here a preliminary exploration suggesting that the addition of BAT does not hurt performance in document classification.

In our setup, we seek to predict document labels y_d from the MAP estimate of a document’s topic distribution, θ_d . Specifically, we classify the newsgroup to which a document was posted for the 20 newsgroups data (e.g., `talk.politics.misc`) and a binary sentiment label for the IMDb review data. We train a random forest classifier using default parameters from `scikit-learn` (Pedregosa et al., 2011) and report the accuracies in Table A.3 (averaged across 5 runs).

Much like other work that is aimed at topic coherence rather than their downstream use in supervised models (Nan et al., 2019), we find that our method has little impact on predictive performance. While it is possible that improvements may be obtained by specifically tuning models for classification, or by integrating BAT into model variations that combine lexical and topic representations (e.g. Nguyen et al., 2013), we leave this to future work.

	K	SCHOLAR	+BAT
20ng	50	0.676 (0.003)	0.669 (0.005)
	200	0.683 (0.002)	0.679 (0.004)
IMDb	50	0.829 (0.003)	0.823 (0.011)
	200	0.805 (0.003)	0.814 (0.004)

Table A.3: Random forest classification accuracy on 20ng and IMDb datasets, using topic estimates from SCHOLAR and SCHOLAR + BAT.

²The entire contents of Appendix A.2 are directly taken from Hoyle et al. (2020).

A.3 Using BAT with DVAE

We further illustrate our method’s modularity by applying BAT to our own reimplementation of DVAE (Burkhardt and Kramer, 2019).³ In contrast to the author’s primary implementation, which estimates the model with rejection sampling variational inference (used in Section 2.1.3), we reimplemented DVAE, approximating the Dirichlet gradient via pathwise derivatives (Jankowiak and Obermeyer, 2018b), similar to Burkhardt and Kramer (2019)’s alternative model variant using implicit gradients.

Our reimplementation shows baseline behavior substantially similar to the author’s implementation. In the course of our experimentation, we noted a degeneracy in this model, in which high NPMI is achieved but at the cost of redundant topics. This failure mode is well-established, but as discussed in Appendix A.3.1, we find the measures proposed to diagnose topic diversity (including those proposed by Burkhardt and Kramer, 2019; Nan et al., 2019) to be problematic. Rather than use these metrics, therefore, we took a coarse but simple approach and filtered out any models that yielded more than one pair of identical topics, averaged across five runs (defined as having two topics with the same set of top-10 words). This filtering eliminated many hyperparameter settings, leading us to believe that DVAE is not robust to this problem.

Ultimately, we find that applying BAT to DVAE does not hurt, and also does not help appreciably (Table A.4). In addition, when applying the above filtering criterion to our main SCHOLAR and SCHOLAR + BAT models, we still obtain the positive results reported in Table A.1.⁴

³We appreciate a reviewer’s suggestion that we add a + BAT comparison for DVAE. The entire contents of Appendix A.3 are directly taken from Hoyle et al. (2020).

⁴For $K = 50$. The single-pair threshold proves too restrictive for the $K = 200$ case, where no

	20NG	Wiki	IMDb
DVAE	0.376 (0.004)	0.517 (0.006)	0.169 (0.007)
+ BAT	0.401 (0.005)	0.515 (0.007)	0.169 (0.006)

Table A.4: Mean development set NPMI (s.d.) across 5 runs for DVAE (Burkhardt and Kramer, 2019) and DVAE + BAT for $K = 50$.

A.3.1 Topic Diversity

Burkhardt and Kramer (2019) found a degeneracy in some topic models, wherein a single topic will be repeated more than once with slightly varying terms (e.g., several Dadaism topics).⁵ Burkhardt and Kramer (2019) and others (Nan et al., 2019; Dieng et al., 2020) have independently proposed related metrics to quantify the problem, but the literature has not converged on a solution. In contrast to NPMI, we are not aware of any work that assesses the validity of such metrics with respect to human judgments.

Moreover, all these proposals suffer from a common problem: because they are global measures of word overlap, they fail to account for *how* words are repeated across topics. For instance, Topic Uniqueness (Nan et al., 2019) is identical regardless of whether all of a topic’s top words are repeated in a single second topic, or individual top words from that topic are repeated in several other topics. In addition, the measures inappropriately penalize partially-related topics.

They also penalize polysemy—and, more generally, the contextual flexibility of word meanings. One of the key *advantages* of latent topics, compared to surface lexical summaries, is that the same word can contribute differently to an understanding of what different topics are about. As a real example from our experience,

hyperparameter settings pass the threshold. Increasing the tolerance to a maximum of 5 redundant pairs with $K = 200$ leads to a somewhat lower average NPMI overall, but the same directional improvement, i.e. SCHOLAR + BAT yields a significantly higher NPMI than SCHOLAR.

⁵The entire contents of Appendix A.3.1 are directly taken from Hoyle et al. (2020).

in modeling a set of documents related to paid family and medical leave, words like *parent*, *mother*, and *father* are prominent in one topic related to parental leave when a child is born (accompanying other terms like *newborn* and *maternity leave*) and also in another topic related to taking leave to care for family members, including elderly parents (accompanying other terms like *elderly* and *aging*). The fact that topic models permit a word like *parent* to be prominent in both of these clearly distinct topics, emphasizing two different aspects of the word relative to the collection as a whole (being a parent taking care of children, being a child taking care of parents), is a feature, not a bug. We consider the question of topic diversity an important direction for future work.

A.4 Experimental Procedures

In this section, we first provide details of our hyperparameters and tuning procedures, then turn to our computing infrastructure and the rough runtime of the SCHOLAR model.

A.4.1 Hyperparameter Tuning and Optimal Values

We used well-tuned baselines to establish thresholds for performance on NPMI (following the reported hyperparameters in [Card et al., 2018](#); [Burkhardt and Kramer, 2019](#); [Nan et al., 2019](#)).⁶ While developing our model, we performed a coarse-grained initial hyperparameter sweep to identify ranges that were not beating the threshold, and decided to exclude those ranges when performing a full grid search. We report the hyperparameter ranges used in this search, along with their optimal values (as determined by development set NPMI), in Tables [A.5](#) to [A.9](#). These

⁶The entire contents of Appendix [A.4.1](#) are directly taken from [Hoyle et al. \(2020\)](#).

Dataset: 20NG		k = 50		k = 200	
	Values Tried	SCHOLAR (optimal values)	SCHOLAR + BAT (optimal values)	SCHOLAR (optimal values)	SCHOLAR + BAT (optimal values)
lr	0.002*	0.002	0.002	0.002	0.002
α	1.0*	1.0	1.0	1.0	1.0
λ	{0.25, 0.5, 0.75, 0.95, 0.99, 0.999}	-	0.75	-	0.99
T	{1.0, 2.0, 3.0, 5.0}	-	2.0	-	5.0

Table A.5: Hyperparameter ranges and optimal values (as determined by development set NPMI) for SCHOLAR and SCHOLAR+ BAT, on the **20NG** dataset. lr is the learning rate, α is the hyperparameter for the logistic normal prior, λ is the weight on the teacher model logits from Eq. (2.4), and T is the softmax temperature from Eq. (2.4). Other hyperparameter values (which can be accessed in our code base) which were kept at their default values are not reported here. Values marked with the * are also kept at their default values per the base SCHOLAR model (<https://github.com/dallascard/scholar>). All different sweeps in the grid search were run for **500 epochs** with a **batch size = 200**.

produced the final set of results (Tables 2.2, 2.3, A.1 and A.2).

For the DISTILBERT training, we use the default hyperparameter settings for the bert-base-uncased model (Wolf et al., 2020). Our code is a modified version of the MM-IMDB multimodal sequence classification code from the same codebase as DISTILBERT, and we use all default hyperparameter settings specified there.⁷ We train for 7500 steps for 20ng, and 17000 steps for Wiki and IMDB (this corresponds to convergence on development-set perplexity).

A.4.2 Computing Infrastructure and Runtime

For the full hyperparameter sweep, we used an Amazon Web Services ParallelCluster (<https://github.com/aws/aws-parallelcluster>) with 40 nodes of g4dn.xlarge instances (consisting of Nvidia T4 GPUs with 16 GB RAM), which ran for about 5 days.⁸ For initial experimentation, we used a SLURM cluster with a mix of consumer-grade Nvidia GPUs (e.g., 1080, 2080).

⁷<https://github.com/huggingface/transformers/tree/master/examples/contrib/mm-imdb>

⁸The entire contents of Appendix A.4.2 are directly taken from Hoyle et al. (2020).

Dataset: Wiki		k = 50		k = 200	
	Values Tried	SCHOLAR (optimal values)	SCHOLAR + BAT (optimal values)	SCHOLAR (optimal values)	SCHOLAR + BAT (optimal values)
lr	{0.001, 0.002, 0.005}	0.001	0.001	0.002	0.005
α	{0.0005, 0.00075, 0.001, 0.005, 0.01}	0.01	0.00075	0.0005	0.001
anneal	{0.25, 0.5, 0.75}	0.25	0.5	0.25	0.5
λ	{0.4, 0.5, 0.6, 0.7, 0.75, 0.8}	-	0.75	-	0.75
T	{1.0, 2.0}	-	1.0	-	1.0
clipping	{1.0, 1.5, 2.0}	-	2.0	-	1.5

Table A.6: Hyperparameter ranges and optimal values (as determined by development set NPMI) for SCHOLAR and SCHOLAR+ BAT, on the **Wiki** dataset. **lr** is the learning rate, **α** is the hyperparameter for the logistic normal prior, **anneal** controls the annealing (as explained in Appendix B in Card et al. (2018)), **λ** is the weight on the teacher model logits from Eq. (2.4), **T** is the softmax temperature from Eq. (2.4), and **clipping** controls how much of the logit distribution to clip (Section 2.1.1). Other hyperparameter values (which can be accessed in our code base) which were kept at their default values are not reported here. All different sweeps in the grid search were run for **500 epochs** with a **batch size = 500**.

Dataset: IMDb		k = 50		k = 200	
	Values Tried	SCHOLAR (optimal values)	SCHOLAR + BAT (optimal values)	SCHOLAR (optimal values)	SCHOLAR + BAT (optimal values)
lr	0.002*	0.002	0.002	0.002	0.002
α	{0.01, 0.1, 0.5, 1.0}	0.5	0.5	0.1	0.1
anneal	{0.25, 0.5, 0.75}	0.25	0.25	0.25	0.5
λ	{0.25, 0.5, 0.75, 0.99}	-	0.5	-	0.99
T	{1.0, 2.0}	-	1.0	-	1.0
clipping	{0.0, 1.0, 10.0}	-	10.0	-	0.0

Table A.7: Hyperparameter ranges and optimal values (as determined by development set NPMI) for SCHOLAR and SCHOLAR+ BAT, on the **IMDb** dataset. **lr** is the learning rate, **α** is the hyperparameter for the logistic normal prior, **anneal** controls the annealing (as explained in Appendix B in Card et al. (2018)), **λ** is the weight on the teacher model logits from Eq. (2.4), **T** is the softmax temperature from Eq. (2.4), and **clipping** controls how much of the logit distribution to clip (Section 2.1.1). Other hyperparameter values (which can be accessed in our code base) which were kept at their default values are not reported here. Values marked with the * are also kept at their default values per the base SCHOLAR model (<https://github.com/dallascard/scholar>). All different sweeps in the grid search were run for **500 epochs** with a **batch size = 200**.

(Dataset: 20NG)			
	Values Tried	W-LDA (optimal values)	W-LDA + BAT (optimal values)
lr	{0.002}	0.002	0.002
α	{0.1, 1.0}	0.1	0.1
λ	{0.75, 0.99}	-	0.75
T	{1.0, 2.0}	-	1.0
(Dataset: Wiki)			
lr	{0.001}	0.001	0.001
α	{0.01, 0.1}	0.1	0.1
λ	{0.25, 0.75}	-	0.25
T	{1.0, 2.0, 5.0}	-	2.0
clipping	{1.0, 2.0}	-	1.0
(Dataset: IMDb)			
lr	{0.002}	0.002	0.002
α	{0.1}	0.1	0.1
λ	{0.75}	-	0.75
T	{1.0}	-	1.0

Table A.8: Hyperparameter ranges and optimal values (as determined by development set NPMI) for W-LDA and W-LDA+ BAT, on all three datasets. **lr** is the learning rate, **α** is the hyperparameter for the dirichlet prior, **λ** is the weight on the teacher model logits from Eq. (2.4), **T** is the softmax temperature from Eq. (2.4), and **clipping** controls how much of the logit distribution to clip (Section 2.1.1). Other hyperparameter values (which can be accessed in our codebase) which were kept at their default values in the original baseline code are not reported here (also see Nan et al. (2019) and <https://github.com/aws-labs/w-l-da/>). Values marked with the * are also kept at their default values. All different sweeps in the grid search were run for **500 epochs** and noise parameter = 0.5 (see Nan et al. (2019)). For 20NG and IMDb, we used **batch size = 200**, and for Wiki, we used **batch size = 360**.

	k = 50			k = 200		
	20NG	Wiki	IMDb	20NG	Wiki	IMDb
Optimal Dirichlet Prior	0.6		0.2	0.6		0.2

Table A.9: For DVAE, we tried four values for the Dirichlet Prior (as per the values tried by the authors in [Burkhardt and Kramer \(2019\)](#)) - $\{0.01, 0.1, 0.2, 0.6\}$ and report the optimal values corresponding to the dev set results (Table 2.2) and test set results (Table A.1) in this table. Within the model variations available in the codebase for DVAE (<https://github.com/sophieburkhardt/dirichlet-vae-topic-models>) we use the Dirichlet VAE based on RSVI which is shown to give the highest NPMI scores in [Burkhardt and Kramer \(2019\)](#).

In terms of runtime, SCHOLAR) and our own SCHOLAR+ BAT are equal and this is true for any of our baseline model augmented with BAT. It is important to note that the overhead in terms of the overall runtime comes only from training the DISTILBERT encoder on the full dataset first and inference time for obtaining the logits after training. Thus, users should keep in mind the initial step of training and inferring teacher model logits and saving them; once that is done for the dataset, our model does not add to the runtime. We show the comparison between the full runtimes, including the initial step, in Fig. A.1.

A.5 Changes to W-LDA

In Fig. A.2, we show the changes to the W-LDA model necessary to accommodate our method.⁹ Ignoring the code to load & clip the logits, also constituting a minor change, we introduce about a dozen lines.

⁹The entire contents of Appendix A.5 are directly taken from [Hoyle et al. \(2020\)](#).

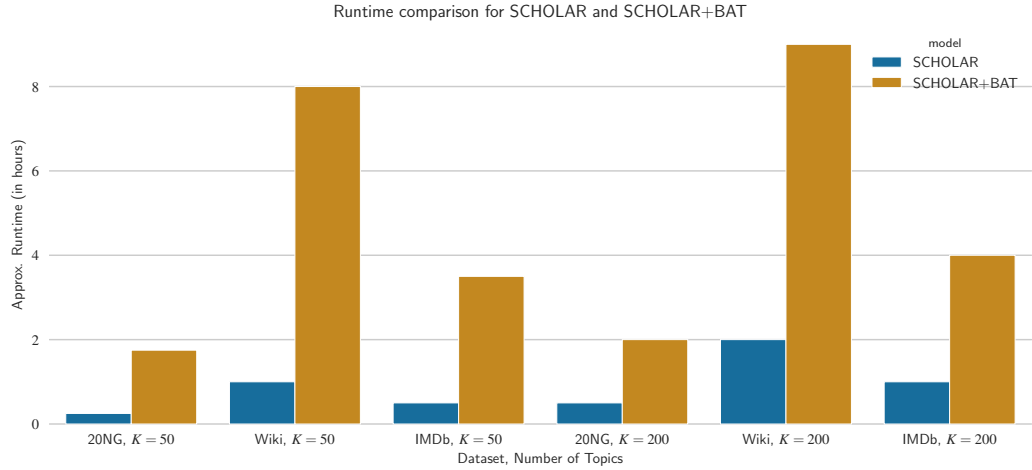


Figure A.1: Runtime comparison for SCHOLAR) and our own SCHOLAR+ BAT - Note that the overhead due to BAT is only a due to the training and inference time required to obtain the DISTILBERT encoder logits on the full dataset first, and once the teacher logits are available, the run time of both models is the same. We depict the full approximate time (in hours) including this initial overhead in case of BAT.

A.6 Impact of BAT on Individual Topics: Aligned Topic Pair Examples

For each corpus (20NG, Wiki, and IMDb), a single comparison of base and BAT-augmented (SCHOLAR vs. SCHOLAR + BAT) 50-topic models was selected randomly, from the five runs used in computing average performance in Fig. 2.4.¹⁰

For each of those pairs of models, we then randomly selected 15 aligned topic pairs from that set of 50 to include in the tables below. Specifically, a full set of 50 topic pairs was partitioned according to JS divergence into the 10 most similar pairs, the next 10 most similar, and so forth, for a total of five “brackets” of topic alignment quality. Three topic pairs were then selected at random from each bracket, hence 15 pairs in all, in order to yield a fair picture of what pairs look like at various

¹⁰The entire contents of Appendix A.6 are directly taken from Hoyle et al. (2020).


```

### In `compute_op.py`

## Retrieve BERT logits
docs = self.data.get_documents(key='train')
if self.args['use_kd']:
    split_on = docs.shape[1] // 2
    docs, bert_logits = docs[:, :split_on], docs[:, split_on:]
    t = self.args['kd_softmax_temp']
    kd_docs = nd.softmax(bert_logits / t) * nd.sum(docs, axis=1, keepdims=True)

# [... unchanged lines ...]

## Compute loss
with autograd.record():
    # [... unchanged lines ...]
    if self.args['use_kd']:
        kd_logits = nd.log_softmax(x_reconstruction_u / t)
        logits = nd.log_softmax(x_reconstruction_u)

        kd_loss_reconstruction = nd.mean(nd.sum(- kd_docs * kd_logits, axis=1))
        loss_reconstruction = nd.mean(nd.sum(- docs * logits, axis=1))

        loss_total = self.args['recon_alpha'] * (
            self.args['kd_loss_alpha'] * t * t * (kd_loss_reconstruction) +
            (1 - self.args['kd_loss_alpha']) * loss_reconstruction
        )
    else:
        # [... unchanged lines ...]

```

Figure A.2: Modified portions of W-LDA model to accommodate BAT. We omit definitions of additional command-line arguments and data loading, but they are similarly brief.

Pair #	SCHOLAR vs SCHOLAR + BAT (NPML, Top 10 Topic Words)	JS Divergence
1	SCHOLAR: (0.399, 'sin eternal lord heaven pray christ prayer jesus god hell') SCHOLAR + BAT: (0.394, 'eternal god hell sin heaven christ jesus christianity faith life')	0.0287
4	SCHOLAR: (0.3512, 'score goal puck penalty season shot tie pitch game defensive') SCHOLAR + BAT: (0.3838, 'score goal season game puck shot leafs penalty play playoff')	0.0345
8	SCHOLAR: (0.4307, 'doctrine church catholic scripture spirit biblical revelation bible resurrection christ') SCHOLAR + BAT: (0.4454, 'biblical church bible scripture doctrine catholic interpretation passage teaching jesus')	0.0417
11	SCHOLAR: (0.7109, 'turks armenian genocide jews mountain armenians turkish proceed nazi armenia') SCHOLAR + BAT: (0.7297, 'turks genocide turkish armenian armenia armenians massacre turkey proceed muslim')	0.0425
15	SCHOLAR: (0.2626, 'cryptography security network privacy mailing internet mail encrypt anonymous user') SCHOLAR + BAT: (0.289, 'anonymous mail network privacy internet security cryptography encrypt electronic ftp')	0.0479
17	SCHOLAR: (0.3501, 'rider bike ride helmet motorcycle dog bmw dod honda seat') SCHOLAR + BAT: (0.3843, 'helmet bike rider ride dog motorcycle dod rear honda bmw')	0.0498
22	SCHOLAR: (0.307, 'voltage circuit amp heat battery electronics frequency signal audio ac') SCHOLAR + BAT: (0.3236, 'circuit voltage amp wire audio wiring signal outlet input pin')	0.0641
27	SCHOLAR: (0.4018, 'passage verse jesus biblical resurrection scripture translation interpretation bible prophet') SCHOLAR + BAT: (0.5262, 'jesus christ lord sin heaven resurrection holy mary father son')	0.071
28	SCHOLAR: (0.2469, 'nt printer windows microsoft mac unix postscript pc os print') SCHOLAR + BAT: (0.2786, 'font color image format printer display pixel graphic postscript directory')	0.0729
31	SCHOLAR: (0.2109, 'crash backup gateway disk windows install memory boot floppy cache') SCHOLAR + BAT: (0.3141, 'disk floppy dos scsi ram cache controller isa swap windows')	0.0864
35	SCHOLAR: (0.2589, 'scientific science disease medicine treatment energy observe observation patient scientist') SCHOLAR + BAT: (0.2705, 'science morality objective scientific moral existence observation universe definition theory')	0.093
40	SCHOLAR: (0.1252, 'interested kit sale advance email address australia thanks april mail') SCHOLAR + BAT: (0.206, 'mail email mailing address list thanks interested fax please send')	0.1173
41	SCHOLAR: (0.2842, 'insurance tax hospital coverage health pay canadian kid care economy') SCHOLAR + BAT: (0.2354, 'dealer car price insurance buy pay sell money honda ford')	0.1319
45	SCHOLAR: (0.3165, 'waco clinton president bush senate batf tax fbi compound vote') SCHOLAR + BAT: (0.5144, 'nsa crypto clipper escrow wiretap secure encryption chip warrant scheme')	0.1791
48	SCHOLAR: (0.2329, 'screen mouse monitor printer inch resolution tube apple font print') SCHOLAR + BAT: (0.275, 'heat fuel tube cool detector radar gas nuclear hole cold')	0.2527

Table A.10: Fifteen aligned topic pairs from the 20NG dataset.

qualities of topic alignment.

In the tables below (Tables A.10 to A.12), we present pairs sorted from best to worst alignment quality. Recall that for NPML, higher is better, and for JS divergence, lower score indicates a higher quality match (or alignment) for the topic pair.

A.7 List of Neural Topic Modeling Works used in our Meta-Analysis

In Table A.13, we report the forty publications used in our meta-analysis (Section 2.2.1), which are sourced from a survey of neural topic models (Zhao et al., 2021a).¹¹

¹¹The entire contents of Appendix A.7 are directly taken from Hoyle et al. (2021).

Pair #	SCHOLAR vs SCHOLAR + BAT (NPML, Top 10 Topic Words)	JS Divergence
1	SCHOLAR: (0.5804, 'prognosis protein symptom intravenous diagnosis syndrome medication abnormality infection dysfunction') SCHOLAR + BAT: (0.5464, 'abnormality prognosis intravenous receptor syndrome antibiotic inflammation diagnosis mutation dos')	0.163
4	SCHOLAR: (0.6036, 'parsec brightest orbiting astronomer planetary brightness luminosity jupiter constellation orbit') SCHOLAR + BAT: (0.586, 'orbiting habitable gliese planetary extrasolar parsec brightness luminosity orbital jupiter')	0.1787
8	SCHOLAR: (0.4432, 'lap peloton uci breakaway sprint ferrari bmc tyre podium sauber') SCHOLAR + BAT: (0.4902, 'lap sprint podium finisher quickest uci mclaren ferrari peloton rosberg')	0.1879
11	SCHOLAR: (0.5662, 'ny renumbering cr realigned intersects intersecting hamlet concurrency routing truncated') SCHOLAR + BAT: (0.5888, 'ny intersects renumbering intersecting realigned cr concurrency routing intersection hamlet')	0.1989
15	SCHOLAR: (0.4866, 'byzantine caliphate ibn caliph byzantium abbasid thrace constantinople vassal umayyad') SCHOLAR + BAT: (0.4686, 'byzantium thrace caliphate nikephoros antioch byzantine envoy umayyad principality constantinople')	0.2076
17	SCHOLAR: (0.4944, 'gubernatorial kentucky reelection republican democrat frankfort candidacy legislator congressman caucus') SCHOLAR + BAT: (0.494, 'gubernatorial reelection legislator congressman candidacy caucus whig democrat kentucky veto')	0.2211
22	SCHOLAR: (0.4069, 'electrification electrified locomotive train nok railway freight oslo commuter nsb') SCHOLAR + BAT: (0.3567, 'nok electrified electrification commuter oslo tramway freight livery bergen locomotive')	0.2391
27	SCHOLAR: (0.4187, 'gatehouse chancel nave stonework anglesey castle demography domesday storey vaulted') SCHOLAR + BAT: (0.4035, 'domesday demography cheshire gatehouse storey borough manor priory mersey avon')	0.2564
28	SCHOLAR: (0.4041, 'frigate brig convoy hm torpedoed rigging destroyer sailed sighted starboard') SCHOLAR + BAT: (0.4126, 'brig frigate privateer rigging schooner sloop corvette sighted indiaman brest')	0.2617
31	SCHOLAR: (0.2876, 'raaf battalion aircrew beachhead moresby amberley brigade usaaf dso jagdgeschwader') SCHOLAR + BAT: (0.5148, 'platoon counterattack bridgehead divisional battalion mortar perimeter brigade beachhead regimental')	0.2651
35	SCHOLAR: (0.3361, 'thanouser filmfare bollywood filmography directorial kumar telugu starred biopic hindi') SCHOLAR + BAT: (0.5322, 'kumar bollywood directorial filmography telugu filmfare prasad malayalam bachchan hindi')	0.2888
40	SCHOLAR: (0.7394, 'batsman wicket bowled bowler bowling wisden cricketer selector inning crease') SCHOLAR + BAT: (0.761, 'bowled wisden selector batsman bowler wicket cricketer crease spinner mcc')	0.3045
41	SCHOLAR: (0.4571, 'statute constitutionality plaintiff unconstitutional defendant judicial appellate amendment jurisdiction judiciary') SCHOLAR + BAT: (0.4569, 'prosecutor prosecution investigator testified testimony conviction convicted verdict sentenced pleaded')	0.3137
45	SCHOLAR: (0.4178, 'edda mahabharata scripture purana goddess poem poetic shiva prose devotional') SCHOLAR + BAT: (0.3379, 'northumbria inscription kingship deity shrine annals worshipped attested buddha vassal')	0.3658
48	SCHOLAR: (0.5286, 'cavalry grenadier flank bridgehead infantry bayonet brigade artillery regiment repulsed') SCHOLAR + BAT: (0.3652, 'dso despatch raaf gallantry adjutant instructor aviator canberra airman citation')	0.544

Table A.11: Fifteen aligned topic pairs from the Wiki dataset.

Pair #	SCHOLAR vs SCHOLAR + BAT (NPML, Top 10 Topic Words)	JS Divergence
1	SCHOLAR: (0.2333, 'scientist monster cgi alien creature scientists attack bullets aliens sci') SCHOLAR + BAT: (0.2636, 'scientist alien creature monster aliens computer cgi space giant scientists')	0.0273
4	SCHOLAR: (0.165, 'vhs copy remember dvd ago tape saw video years loved') SCHOLAR + BAT: (0.1844, 'vhs copy tape remember dvd bought ago saw video available')	0.0327
8	SCHOLAR: (0.1146, 'kids kid dad parents mom christmas decides dies santa guy') SCHOLAR + BAT: (0.118, 'dad mom kids parents kid uncle decides christmas dies cat')	0.0379
11	SCHOLAR: (0.1968, 'adaptation version novel bbc versions jane kenneth handsome adaptations faithful') SCHOLAR + BAT: (0.2181, 'adaptation novel book read books faithful bbc version versions novels')	0.0383
15	SCHOLAR: (0.2758, 'show episodes episode shows abc season aired sitcom television seasons') SCHOLAR + BAT: (0.2678, 'seasons episodes show aired episode abc sitcom season television network')	0.0416
17	SCHOLAR: (0.0863, 'fails wooden lacks unconvincing shallow contrived wretched embarrassing thin embarrassment') SCHOLAR + BAT: (0.1047, 'lacks pacing fails contrived flat irritating lacking chemistry unconvincing uninteresting')	0.0424
22	SCHOLAR: (0.174, 'documentary footage interviews music documentaries disc dvd musicians extras insight') SCHOLAR + BAT: (0.1796, 'footage available documentary release dvd print interviews vhs subtitles audio')	0.0459
27	SCHOLAR: (0.1532, 'sheriff car town decides husband killer police investigate chase security') SCHOLAR + BAT: (0.2504, 'murder murdered detective killer murderer police murders suspects secretary serial')	0.0531
28	SCHOLAR: (0.3054, 'christian religious god religion christ faith church jesus beliefs truth') SCHOLAR + BAT: (0.1027, 'filmmaker intellectual filmmakers pretentious artistic subject content sake context claim')	0.0539
31	SCHOLAR: (0.1136, 'gags school rock band cartoons record boys principal radio metal') SCHOLAR + BAT: (0.3144, 'songs musical singing sing dancing singer concert song numbers dance')	0.0556
35	SCHOLAR: (0.0813, 'development seemed boring predictable weak explanation slow potential interesting suspense') SCHOLAR + BAT: (0.092, 'hour asleep minutes seemed sounded sat felt rented waste confusing')	0.0616
40	SCHOLAR: (0.1821, 'noir murder detective gritty crime cop thriller tough clint veteran') SCHOLAR + BAT: (0.1027, 'cop dennis sheriff gangster boss agent villain hopper action chases')	0.0679
41	SCHOLAR: (0.095, 'porn cops girls random camera amateurish tedious amateur screaming chick') SCHOLAR + BAT: (0.1548, 'kills killed killer screaming killing kill boyfriend woods walks dies')	0.0803
45	SCHOLAR: (0.1884, 'planet wars sci graphics space science game robot fiction weapons') SCHOLAR + BAT: (0.0959, 'action development fighting sequences visuals realistic epic fight battles cool')	0.0925
48	SCHOLAR: (0.1732, 'book read books novel adaptation author reading disappointed adapted translation') SCHOLAR + BAT: (0.1063, 'liked overall surprised disappointed enjoyed pleasantly pretty expectations seemed expecting')	0.1196

Table A.12: Fifteen aligned topic pairs from the IMDB dataset.

Source	Human Evals?	Perplexity	Coherence	Implementation Specified	Ref. Corpus Specified ?	Consistent Preproc?	Hparam search?	χ^2 run / err. bars?	LDA Implementation?	Baseline w / in 2 yr?
Bianchi et al. (2021)	No	No	NPMI, Embed-sim	None	Internal, External-GoogleNews	Yes	No	Yes	Variational	No
Zhao et al. (2021b)	No	No	NPMI	Palmetto	No	Unclear	No	Yes	N/A	Yes
Feng et al. (2022)	No	Yes	NPMI	None	No	Yes	No	No	N/A	Yes
Hoyle et al. (2020)	No	No	NPMI	In paper	External NYT, Internal	No	Yes	Yes	N/A	Yes
Hu et al. (2020b)	No	No	C_p , C_v , NPMI	Palmetto	External WIKI	No	Likely no	No	Sampling	Yes
Isonuma et al. (2020)	No	Yes	NPMI	None	No, likely external	Unclear	No	No	Sampling	No
Joo et al. (2020)	No	Yes	NPMI	None	No, likely internal	Unclear	Likely yes	Yes	N/A	Yes
Lin et al. (2020)	No	Yes	NPMI	None	No, likely internal	Unclear	Yes	Yes	N/A	Yes
Ning et al. (2020)	No	Yes	NPMI	Lau github	No	Yes	Likely no	Yes	Variational	No
Panwar et al. (2020)	No	No	NPMI	Lau github	No	Yes	Likely no	No	Sampling	Yes
Rezaee and Ferraro (2020)	No	No	N/A	N/A	N/A	Yes	Likely no	Yes	Variational	No
Thompson and Mimmo (2020)	No	No	Coherence, PMI	In paper	External NYT	No	Likely no	Yes	Sampling	No
Tian et al. (2020)	No	Yes	NPMI	None	No	No	Yes	No	Variational	Yes
Wang et al. (2020a)	No	No	C_p , C_v , NPMI, UCI	Palmetto	No	No	No	No	Sampling	Yes
Wu et al. (2020a)	No	Yes	NPMI	None	No	No	Yes	No	N/A	Yes
Wu et al. (2020b)	No	Yes	C_v	Palmetto	No	Yes	No	No	Unspecified	Yes
Yang et al. (2020a)	No	Yes	Coherence	In paper	No, likely internal	Yes	No	No	Unspecified	No
Zhou et al. (2020)	No	No	NPMI, C_p	Palmetto	External WIKI	No	Likely no	No	Unspecified	Yes
Burkhardt and Kramer (2019)	No	Yes	NPMI	None	No, likely internal	Unclear	Yes	No	Variational	Yes
Dieng et al. (2020)	No	Yes	Coherence	In paper	No, likely internal	Yes	No	No	Unspecified	No
Gui et al. (2019)	No	No	C_v	None	External WIKI	Yes	Likely no	No	Unspecified	Yes
Gupta et al. (2019b)	No	Yes	C_v	Gensim	No, likely internal	Unclear	Likely no	No	N/A	No
Gupta et al. (2019a)	No	Yes	C_v	Gensim	No, likely internal	Unclear	Likely no	No	Sampling	Yes
Lin et al. (2019)	No	Yes	PMI	In paper	No, likely external	Unclear	No	No	Variational	Yes
Liu et al. (2019a)	No	Yes	NPMI	Lau github	No, likely internal	Yes	No	No	Variational	Yes
Nan et al. (2019)	No	No	NPMI	None	No	No	No	No	Sampling	Yes
Wang et al. (2020b)	No	No	C_p , C_v , UCI, NPMI, UMass	Palmetto	No	No	No	No	Unspecified	Yes
Card et al. (2018)	No	Yes	NPMI	In paper	External-gigaword	Yes	Likely yes	No	Sampling	Yes
Ding et al. (2018)	No	Yes	NPMI	Lau github	No, likely external	No	Likely no	No	Sampling	Yes
He et al. (2018)	No	No	Coherence	None	No, likely internal	Yes	No	No	N/A	Yes
Peng et al. (2018)	No	Yes	N/A	N/A	N/A	Yes	Likely no	No	Variational	Yes
Silveira et al. (2018)	No	Yes	NPMI	Lau github	Internal	Yes	No	Yes	N/A	Yes
Zhang et al. (2018)	No	Yes	N/A	N/A	N/A	Unclear	Likely no	No	N/A	Yes
Zhao et al. (2018)	No	Yes	NPMI	Palmetto	External WIKI	Unclear	No	Yes	N/A	Yes
Zhu et al. (2018)	No	No	Coherence	None	No, likely internal	Yes	Likely no	No	Variational	Yes
Jung and In Choi (2017)	No	Yes	NPMI, PMI, UMass	None	No	Yes	No	No	Sampling	Yes
Miao et al. (2017)	No	Yes	NPMI	In paper	No	No	Likely no	No	Variational	Yes
Srivastava and Sutton (2017)	No	Yes	NPMI	None	No	Yes	No	No	Sampling	Yes
Miao et al. (2016)	No	Yes	N/A	N/A	N/A	Yes	Likely no	No	Unspecified	Yes
Nguyen et al. (2015a)	No	No	NPMI	Lau github	External WIKI	Yes	No	Yes	Sampling	No

Table A.13: Papers used in meta-analysis, Section 2.2.1

A.8 Preprocessing Details

Our steps are delineated in our implementation,¹² but we list our choices here for easy reference.¹³ Corpus statistics are provided in Table 2.8. We use the default `en-core-web-sm` spaCy model (Honnibal et al., 2020), version 3.0.5, throughout.

Document processing

- We do not process documents with fewer than 25 whitespace-separated tokens.
- Following processing (e.g., stopword removal), we remove documents with fewer than five tokens.
- We truncate documents to 5,000 whitespace-separated tokens for NYT and to 19,000 for WIKI (in both cases affecting less than 0.15% of documents).

Vocabulary creation

- We tokenize using spaCy.
- We lowercase terms.
- We *do not* lemmatize.
- We detect noun entities with spaCy, keeping only the `ORG`, `PERSON`, `FACILITY`, `GPE`, and `LOC` types, joining constituent tokens with an underscore (e.g., “New York City” → `new_york_city`).

Vocabulary filtering

¹²github.com/ahoho/topics

¹³The entire contents of Appendix A.8 are directly taken from Hoyle et al. (2021).

- The vocabulary is created from the training data. The reference texts used in coherence calculations are processed identically and use the same vocabulary.
- We filter out stopwords using the default spaCy English stopwords list.¹⁴ Stopwords are retained if they are contained within detected noun entities (e.g., “The United States of America” → `united_states_of_america`).
- We filter out tokens with two or fewer characters.
- We retain only tokens that are matched by the regular expression:

$$\wedge [\wedge -] * [a-zA-Z] [\wedge -] * \$$$
- We remove tokens that appear in more than 90% of documents.
- We remove tokens that appear in fewer than $2(0.02|D|)^{1/\log 10}$ documents, where $|D|$ is the corpus size.¹⁵

A.9 Training Details

Expanding Section 2.2.3, we detail the hyperparameter tuning for each of our three topic models, along with other pertinent details about runtimes and compute resources.¹⁶ Scripts used to run the models with all the various hyperparameter configurations are released as part of our code; this section is also included for reference.

¹⁴github.com/explosion/spaCy/blob/v3.0.5/spacy/lang/en/stop_words.py

¹⁵Standard rules-of-thumb for vocabulary pruning, like removing terms that appear in fewer than 0.5% of documents (Denny and Spirling, 2018), ignore the power-law distribution of word frequency Zipf (1949), and hence do not scale to large corpora. To keep vocabulary sizes roughly consistent across datasets, we set the minimum document-frequency for terms as a (power) function of the total corpus size. This has the intuitive appeal of increasing proportional to the order of magnitude of the number of total documents, starting at a minimum document-frequency of 2 for a 50-document corpus and reaching about 110 for a corpus of 500,000.

¹⁶The entire contents of Appendix A.9 are directly taken from Hoyle et al. (2021).

Our general strategy, especially with the neural models, is to select different values around the reported optimal settings in original papers. For all three models, we try two different values for the number of training iterations (G-LDA) or epochs (D-VAE, ETM).

G-LDA We use gensim ([Řehůřek and Sojka, 2010](#)) as a Python wrapper for running Mallet. In Table [A.14a](#), we tune hyperparameters α (topic density parameter) and β (word density parameter) which can be thought of as “smoothing parameters” that reserve some probability for the topics (words) unassigned to a document (topic) thus far. Mallet internally optimizes hyperparameters, and the Optimization Interval controls the frequency of hyperparameter updates, measured in training steps.

D-VAE Our reimplementation of Dirichlet-VAE ([Burkhardt and Kramer, 2019](#)) largely uses the same hyperparameters as reported in that work. As shown in Table [A.14b](#), we vary the prior for the Dirichlet distribution (α), the learning rate (η), the L_1 -regularization constant for the topic-word distribution ($\beta_{reg.}$, not in the original model but inspired by [Eisenstein et al., 2011](#)), the number of epochs to anneal the use of batch normalization in the decoder (γ_{BN} , comes from [Card et al., 2018](#)), and the number of epochs to anneal the KL-divergence term in the loss (γ_{KL}) (it needs to be introduced slowly in the loss function due to the component collapse problem in VAEs ([Bowman et al., 2016](#))).

ETM Following [Dieng et al. \(2020\)](#), we learn skip-gram embeddings on the training corpus using the provided script, which relies on gensim. As shown in Table [A.14c](#), we vary the learning rate (η), the L_2 regularization constant for the Adam ([Kingma and Ba, 2015](#)) optimizer (W_{decay}), and a boolean indicator of whether to anneal the

Model: G-LDA			
α	β	Optim. Interval	#Steps
$\{0.01, 0.05, 0.1, 0.25^{\dagger}, 1.0^*, 5.0\}$	$\{0.01, 0.05^*, 0.1^{\dagger}\}$	$\{0, 10^{\dagger}, 100, 500^*\}$	$\{1000^{\dagger}, 2000^*\}$

(a) Hyperparameter ranges for G-LDA. α is the topic density parameter. β is the word density parameter. Optim. Interval sets the number of iterations between Mallet’s own internal hyperparameter updates. #Steps are training iterations.

Model: D-VAE					
α	η	$\beta_{reg.}$	γ_{BN}	γ_{KL}	#Steps
$\{0.001, 0.01^{*\dagger}, 0.1\}$	$\{0.001, 0.01^{*\dagger}\}$	$\{0.0^*, 0.01, 0.1^{\dagger}, 1.0\}$	$\{0, 1^*, 100, 200^{\dagger}\}$	$\{100^*, 200^{\dagger}\}$	$\{200, 500^{*\dagger}\}$

(b) Hyperparameter ranges for D-VAE. α is the Dirichlet prior. η is the learning rate. $\beta_{reg.}$ is the L_1 -regularization of the topic-word distribution. γ_{BN} and γ_{KL} are the number of epochs to anneal the batch normalization constant and KL divergence term in the loss, respectively. #Steps are training epochs.

Model: ETM			
η	W_{decay}	γ_{η}	#Steps
$\{0.001^*, 0.002, 0.01, 0.02^{*\dagger}\}$	$\{1.2e^{-5^*}, 1.2e^{-6^{\dagger}}, 1.2e^{-7}\}$	$\{0^{*\dagger}, 1\}$	$\{500, 1000^{*\dagger}\}$

(c) Hyperparameter ranges for ETM. η is the learning rate. W_{decay} is the L_2 regularization constant. γ_{η} is an indicator of whether learning rate is annealed. #Steps are training epochs.

Table A.14: Hyperparameter settings for G-LDA, D-VAE, and ETM. *: Best setting for WIKI, †: best setting for NYT; based on NPMI estimated with a 10-token sliding window over the reference corpus.

learning rate (γ_{η}). If annealing is allowed, the learning rate gets divided by 4.0 if the loss on the validation set does not improve for more than 10 epochs, per the default settings of the model (preliminary experiments showed that annealing did not attain higher NPMI).

The runtimes for each of the models on each dataset are in Table A.15. We used AWS ParallelCluster to provide a cloud-computing computing cluster. Neural models ran on NVIDIA T4 GPUs using g4dn.xlarge instances with 16 GiB memory and 4 CPUs.¹⁷ G-LDA (Mallet) ran on CPU only, with m5d.2xlarge instances (with 32 GiB memory, 8 CPUs).¹⁸

¹⁷<https://aws.amazon.com/hpc/parallelcluster/>

¹⁸See <https://aws.amazon.com/ec2/instance-types/> for further details.

	WIKI	NYT
G-LDA	~ 2 minutes	~ 9 minutes
D-VAE	~ 45 minutes	~ 330 minutes
ETM	~ 260 minutes	~ 1300 minutes

Table A.15: Runtimes for the three topic models on each of the two datasets. G-LDA requires CPUs only while the neural models use a single GPU. Compute resources detailed at the end of Appendix A.9.

A.10 Instructions for Crowdworkers

Recruiting participants on Prolific.co for a Qualtrics survey produced results with higher inter-worker agreement than Mechanical Turk, based on a pilot test.¹⁹ Using the Prolific.co platform, we recruited respondents that met the criteria of living in the United States and listing fluency in English. Each respondent was paid through Prolific upon completion of the survey, at a rate corresponding to \$15 an hour. The total amount spent on conducting all the surveys, including our pilot test, was \$2084.91. We used automated scripts to generate separate Qualtrics surveys for each task that contained the topics for evaluation, available in our released code. Each respondent was shown 25% of the questions in each survey; the question selection and answer display order was chosen randomly via the survey configuration on Qualtrics. Figs. 2.6 and 2.7 depict our word intrusion and ratings tasks, respectively. Crowdworkers receive instructions explaining the task (Fig. A.3) and the dataset (Fig. A.4).

This survey asks you to look at lists of words produced by an automatic computer program. For each list, you'll be answering the question: "Which word doesn't belong?"

- You will be shown ten sets of six words.
- For each set, click the word whose meaning or usage is most unlike that of the other words.
- If you feel that multiple words do not belong, choose the one that you feel is most out of place.
- Do not base your decisions on how the word is pronounced or written or its grammatical function. For example, if you saw (apple, apricot, anvils, peach), you would not choose "peach" because it doesn't start with "a", you would not choose "apricot" because it isn't five letters long, and you would not choose "apple" because it ends with a vowel. Ideally, you would choose "anvils" because it is not a fruit.

Here are some examples:

"baby", "crib", "diaper", "beer", "pacifier", "cry"

In this example the word 'beer' is the least related. All of the other words are closely related to each other, and related to infants.

Here is another, harder, example:

"Hard Drive", "motherboard", "video card", "processor", "RAM", "USB key"

While all of these terms are related to a computer, all but one of them are components inside of a computer. The best choice is therefore 'USB key'.

You may not always know all the words and that's okay.

This study should take approximately 10-15 minutes to complete. Your response will be completely anonymous.

(a)

This survey asks you to evaluate lists of words produced by an automatic method.

The computer model we are testing seeks to identify groups of words that are highly related to each other. You will be asked to select how related groups of words are on a 3-point scale.

The rating options are: Not Very Related, Somewhat Related, Very Related.

A helpful question to ask yourself is: "what is this group of words about?" If you can answer easily, then the words are probably related. Here is some guidance on how to apply these ratings and some examples.

Very Related - Most of the words are clearly related to each other, and it would be easy to describe how they are related.

Example: "dog", "cat", "hamster", "rabbit", "snake" (An obvious way to describe the relationship here would be 'Pets')

Example: "brushwork", "canvases", "expressionism", "cubism", "modernism", "curators", "abstract_expressionism", "national_gallery_of_art", "museum", "fossils" (An obvious way to describe this would be "art", even though one or two of the words are not as clearly related to that.)

Somewhat Related - The words are loosely related to each other, but there may be a few ambiguous, generic, or unrelated words

Example: "computer", "video", "new", "plug", "screen", "model" (In this example, some of the words are generic, and seem more closely related than others)

Example: "dog", "ball", "pet", "receipt", "pen" (In this example, some of the words seem closely related, but not all of them)

Not Very Related - The words do not share any obvious relationship to each other. It would be difficult to describe how the words are related to each other.

Example: "dog", "apple", "pencil", "earth", "computer"

This study should take approximately 10-15 minutes to complete. Your response will be completely anonymous.

(b)

Figure A.3: Instructions for (a) word intrusion and (b) ratings

In this survey, the word lists are based on a computer analysis of The New York Times.

The New York Times is an American newspaper featuring articles from 1987 to 2007. Sections from a typical paper include International, National, New York Regional, Business, Technology, and Sports news; features on topics such as Dining, Movies, Travel, and Fashion; there are also obituaries and opinion pieces.

(a)

In this survey, the word lists are based on a computer analysis of Wikipedia.

Wikipedia is an online encyclopedia covering a huge range of topics. Articles can include biographies ("George Washington"), scientific phenomena ("Solar Eclipse"), art pieces ("La Danse"), music ("Amazing Grace"), transportation ("U.S. Route 131"), sports ("1952 winter olympics"), historical events or periods ("Tang Dynasty"), media and pop culture ("The Simpsons Movie"), places ("Yosemite National Park"), plants and animals ("koala"), and warfare ("USS Nevada (BB-36)"), among others.

(b)

Figure A.4: Descriptions for (a) NYTimes and (b) Wikipedia.

A.11 Regression Results

Prior work (e.g., Röder et al., 2015) relates averaged human ratings to automated metrics using either Pearson or Spearman correlations.²⁰ As an alternative that takes into account both the variation in human judgments as well as their numerical type, we estimate logistic and ordered probit regressions on the ratings and intrusion

¹⁹The entire contents of Appendix A.10 are directly taken from Hoyle et al. (2021).

²⁰The entire contents of Appendix A.11 are directly taken from Hoyle et al. (2021).

Ref. Corpus → Train Corpus ↓		NPMI (10-token window)				C_v (110-token window)			
		NYT	WIKI	Train	Val	NYT	WIKI	Train	Val
Intrusion	NYT	2.42	4.16	2.11	1.97	2.50	<u>3.27</u>	2.55	2.40
	WIKI	<u>4.11</u>	<u>5.08</u>	5.45	0.87	2.23	2.79	2.74	0.70
	Concatenated	2.82	4.56	3.18	0.78	2.30	3.05	2.64	0.87
Rating	NYT	1.92	<u>2.08</u>	1.77	1.85	<u>2.55</u>	<u>2.51</u>	2.68	<u>2.59</u>
	WIKI	2.97	<u>4.10</u>	4.29	1.45	2.01	2.82	2.86	0.80
	Concatenated	<u>2.20</u>	2.75	<u>2.52</u>	1.17	<u>2.27</u>	<u>2.60</u>	<u>2.74</u>	1.07

Table A.16: Logistic (intrusion) and ordinal probit (ratings) regression coefficients of automated metrics on human annotations. Underlined values have overlapping 95% confidence intervals with that of the **largest** value in each row.

annotations, respectively. In Table A.16, we report the estimated coefficients for each metric, finding that—on the whole—using the WIKI reference performs best, although the large estimated confidence intervals mitigate the strength of this conclusion.

A.12 Filtering on Term Familiarity

Several topics, particularly those produced by D-VAE, contain terms that are not well-known to annotators (Section 2.2.5).²¹ When a respondent is unfamiliar with a topic’s words, their ratings for that topic may not accurately reflect its true coherence. For example, a mycologist may find the words in the fifth column of Table 2.6 highly related, whereas someone unfamiliar with fungi-related jargon may rate it poorly—indeed, the mean rating for this topic is 2.1 for those unfamiliar with terms and 2.6 for those who are familiar.

Since automated metrics do not take into account a term’s familiarity to humans, we posit that automated metrics should be more predictive of human judgments among respondents who are familiar with topic terms. To test this hypothesis, we

²¹The entire contents of Appendix A.12 are directly taken from Hoyle et al. (2021).

		NPMI (10-token window)				C_v (110-token window)			
	Ref. Corpus → Train Corpus ↓	NYT	WIKI	Train	Val	NYT	WIKI	Train	Val
Intrusion	NYT	0.34	<u>0.51</u>	0.32	0.25	0.44	0.55	0.42	0.38
	WIKI	<u>0.39</u>	<u>0.39</u>	<u>0.40</u>	0.14	<u>0.38</u>	0.40	<u>0.39</u>	0.13
	Concatenated	<u>0.36</u>	<u>0.45</u>	<u>0.36</u>	0.18	0.41	0.48	0.41	0.26
Rating	NYT	0.45	0.59	0.44	0.43	0.51	<u>0.58</u>	<u>0.53</u>	0.52
	WIKI	<u>0.45</u>	0.51	<u>0.51</u>	0.21	<u>0.44</u>	<u>0.51</u>	<u>0.51</u>	0.23
	Concatenated	0.47	0.54	0.47	0.35	<u>0.49</u>	<u>0.53</u>	<u>0.51</u>	0.42

(a) Spearman correlation coefficients between mean human scores and automated metrics, compare to Table 2.9.

		NPMI (10-token window)			C_v (110-token window)		
	Ref. Corpus → Train Corpus ↓	NYT	WIKI	Train	NYT	WIKI	Train
Intrusion	NYT	53 / 55	46 / 50	56 / 52	41 / 34	28 / 27	42 / 33
	WIKI	38 / 76	36 / 77	32 / 76	29 / 37	31 / 43	33 / 39
	Concatenated	54 / 70	41 / 73	51 / 70	41 / 44	29 / 42	41 / 44
Rating	NYT	45 / 49	39 / 53	45 / 47	18 / 27	16 / 24	17 / 25
	WIKI	37 / 73	25 / 74	30 / 70	28 / 31	19 / 33	18 / 27
	Concatenated	45 / 64	38 / 68	42 / 64	26 / 36	21 / 36	27 / 33

(b) False discovery rate (1 – precision, lower is better) and false omission rate of significant model differences when using automated metrics, compared to table 2.10.

		NPMI (10-token window)				C_v (110-token window)			
	Ref. Corpus → Train Corpus ↓	NYT	WIKI	Train	Val	NYT	WIKI	Train	Val
Intrusion	NYT	3.71	7.14	3.04	2.54	3.34	4.54	3.23	2.94
	WIKI	<u>5.87</u>	6.46	6.19	0.85	3.23	3.59	3.39	0.42
	Concatenated	4.24	6.81	4.17	0.94	3.18	4.06	3.30	0.91
Rating	NYT	4.40	5.87	3.85	3.93	3.97	4.44	4.03	3.89
	WIKI	<u>4.84</u>	5.95	<u>5.65</u>	1.33	2.96	3.73	3.69	0.62
	Concatenated	4.49	5.80	4.56	1.78	3.45	3.91	3.81	1.32

(c) Logistic (intrusion) and ordinal probit (ratings) regression coefficients of automated metrics on human annotations, compare to Table A.16.

Table A.17: Tables 2.9, 2.10, and A.16 after removing respondents who report a lack of familiarity with topic words.

Ref. Corpus → Train Corpus ↓		NPMI (10-token window)				C _v (110-token window)			
		NYT	WIKI	Train	Val	NYT	WIKI	Train	Val
Rating (5-pt.)	NYT	0.27	<u>0.37</u>	0.28	<u>0.33</u>	0.29	<u>0.35</u>	<u>0.33</u>	<u>0.35</u>
	WIKI	0.15	0.21	0.29	<u>0.43</u>	0.10	0.16	0.17	<u>0.50</u>
	Concatenated	0.21	0.30	0.28	0.32	0.20	0.26	0.26	<u>0.39</u>

Table A.18: Spearman correlation coefficients between mean human scores for a **five-point** ratings scale (rather than three), compared to table 2.9. Underlined values have overlapping 95% confidence intervals with that of the **largest** value in each row.

re-evaluate the relationships between automated metrics and human judgments *after removing* respondents who state they are not familiar with a topic’s terms (Table A.17). On the whole, results are much clearer than above; NPMI estimated using WIKI reference counts is strongly correlated across tasks and datasets. The false discovery rate is lower overall, although automated metrics still misdiagnose significant results at a rate of one in six in even the best case.

These findings provide further evidence—per our discussion in Section 2.2.6—that future human evaluations of topic models ought to take into account domain expertise and information need.

A.13 Five-point Ratings Scale

Although most prior work uses three-point scales for the ratings task (Fig. 2.7), for comparison we also ask annotators to label the topic topic words with a five-point scale ranging from 1 (“not at all related”) to 5 (“very related”, no labels are given for points 2-4).²² Broadly, we find that values for correlations are reduced relative to the three-point scale (Table A.18). We believe examining this discrepancy is an interesting direction for future work that re-visits human evaluation of topic

²²The entire contents of Appendix A.13 are directly taken from Hoyle et al. (2021).

models.

A.14 Potential Negative Impact

Our work focuses its investigation on data from the English language alone.²³ In this way, it further entrenches English-language primacy in NLP, and more crucially, findings may not translate directly to other languages. We caution the reader against applying claims made in this work to topic modeling on corpora of other languages. It is even possible that one of the tasks designed to elicit human judgment (e.g., word intrusion) may not be amenable for use with other languages.

Concerning topic models more broadly, we note that others question the scholarly value of “distant reading” and the digital humanities in general (Marche, 2012; Allington et al., 2016). Do topic models encourage a passive, disengaged relationship to texts—fomenting conclusions about broad, generic trends rather than idiosyncratic specifics, leading us to miss the trees for the forest? As noted by Schmidt (2012), “topics neither can nor should be studied independently of a deep engagement in the actual word counts that build them.” In this light, topic models can be viewed as an extension of the insidious neoliberal trend toward mass data harvesting that blurs differences between individuals and cultures. Researchers should take care to avoid such elisions when drawing conclusions from model outputs.

²³The entire contents of Appendix A.14 are directly taken from Hoyle et al. (2021).

Appendix B: Motivations behind Communicative Choices: Donor Activity is Associated with US Legislators’ Attention to Political Issues

B.1 Additional Experimental Detail

Using the entire dataset, we use one particular randomly selected fold or training-validation split of the data to first tune various hyperparameters of the logistic regression model in order to select the best settings for each of the speaker models. Specifically, we allow the optimal learning rate (α) to be selected from $\{0.000001, 0.000005, 0.00001, 0.00005\}$ and the weight decay (constant for L_2 regularization — λ) to be selected from $\{0.0, 0.001, 0.005, 0.01\}$. We set the number of training epochs to be 10,000 - and tune a binary (1/0 corresponding to yes/no) hyperparameter (η) to enable early stopping (training to stop if the cross entropy loss does not reduce for a certain number of epochs instead of training for the full set of epochs) or not for each of the model. L_2 regularization and early stopping help prevent the model from overfitting on the training set.

The selected settings for each of the speaker models are provided in Table [B.1](#). The results of the 50-fold cross-validation can then be compared across the legislator representations (Fig. [3.5a](#)).

	PAC	Industry	Category	Committee	State	Party
α	0.00001	0.000005	0.000001	0.00005	0.00005	0.00005
λ	0.01	0.005	0.005	0.001	0.001	0.0
η	1	1	1	1	1	1

Table B.1: Values for optimal hyperparameter settings for the multinomial logistic regression model. α = learning rate; $\lambda = L_2$ regularization constant or the weight decay; η = binary indicator of whether to use early stopping (1) or not (0). The best settings found for PAC were also used for Random-PAC.

	PAC	Industry	Category	Committee	State	Party	Random-PAC
1995-96	(312, 617)	(312, 69)	(312, 215)	(312, 81)	(312, 50)	(312, 3)	(312, 617)
1997-98	(382, 648)	(382, 69)	(382, 218)	(382, 85)	(382, 52)	(382, 3)	(382, 648)
1999-00	(415, 694)	(415, 70)	(415, 225)	(415, 87)	(415, 54)	(415, 3)	(415, 694)
2001-02	(416, 743)	(416, 71)	(416, 228)	(416, 85)	(416, 54)	(416, 3)	(416, 743)
2003-04	(418, 754)	(418, 72)	(418, 230)	(418, 87)	(418, 54)	(418, 3)	(418, 754)
2005-06	(421, 800)	(421, 72)	(421, 235)	(421, 91)	(421, 54)	(421, 3)	(421, 800)
2007-08	(416, 823)	(416, 72)	(416, 241)	(416, 92)	(416, 53)	(416, 2)	(416, 823)
2009-10	(396, 835)	(396, 71)	(396, 243)	(396, 93)	(396, 55)	(396, 2)	(396, 835)
2011-12	(401, 875)	(401, 71)	(401, 247)	(401, 92)	(401, 54)	(401, 2)	(401, 875)
2013-14	(402, 854)	(402, 71)	(402, 247)	(402, 89)	(402, 53)	(402, 2)	(402, 854)
2015-16	(355, 837)	(355, 71)	(355, 247)	(355, 89)	(355, 52)	(355, 2)	(355, 837)
2017-18	(303, 810)	(303, 71)	(303, 246)	(303, 89)	(303, 50)	(303, 2)	(303, 810)

Table B.2: The sizes or dimensionality of each of our legislator attributes or representations for when the models are trained separately for each Congressional cycle. The first value in each dimensionality is the number of speakers selected in that cycle, and the second value is the number of features (for the speaker representation or model indicated in the column title). Note that the y for each of the regression models (averaged topic distribution for speakers) is going to be (number of speakers, 30).

The sizes of various legislator attributes or representations for the model run on the entire dataset were included in Table 3.1; these sizes for models trained on each Congressional cycle are shown in Table B.2. 30-fold cross-validation is used. The hyperparameters provided for the attributes or representations in Table B.1 are also used for the respective representation for per-cycle model training.

B.2 Additional Results

The training set results corresponding to Fig. 3.5a and Fig. 3.5b are shown in Fig. B.1 and Fig. B.2 respectively. The cross-validation results for per-cycle models

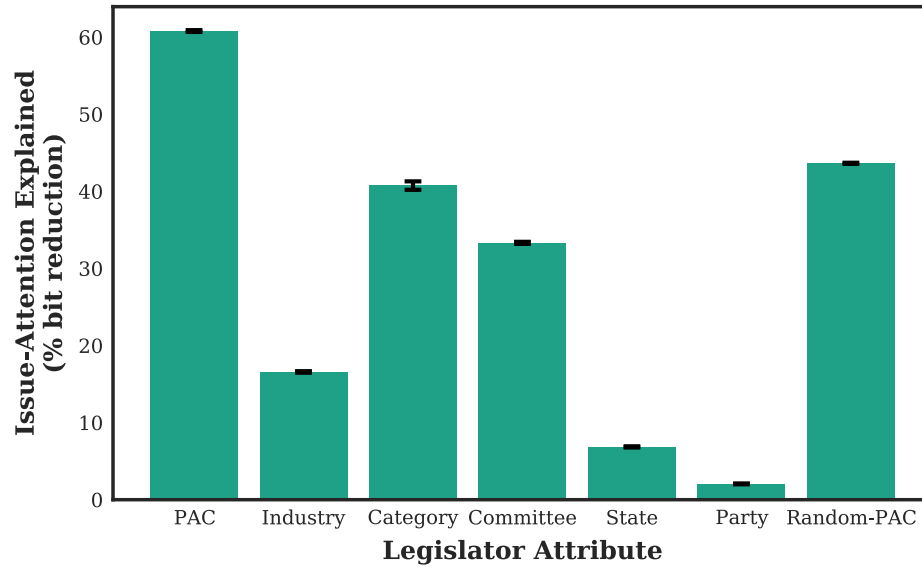


Figure B.1: Training set results when modeling the entire dataset and predicting legislators' issue-attention. The corresponding held-out set results are shown in Fig. 3.5a.

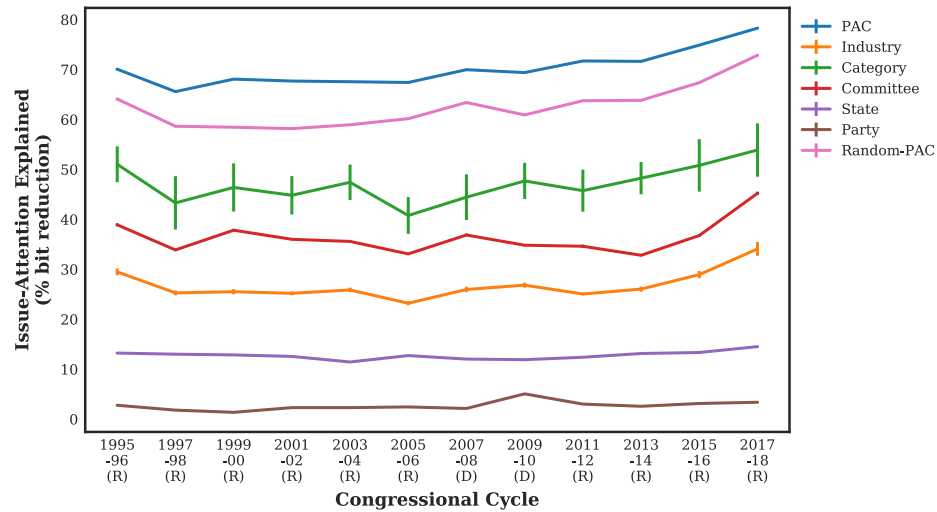


Figure B.2: Training set results when modeling each congressional session separately and predicting legislators' issue-attention. The corresponding held-out set results are shown in Fig. 3.5b.

with error bars corresponding to the cross-validation results without error bars (Fig. 3.5b) are shown in Fig. B.3.

Based on labels provided by experts, we remove certain non-meaningful and non-policy-related topics and retrain the machine learning model. Topics assigned

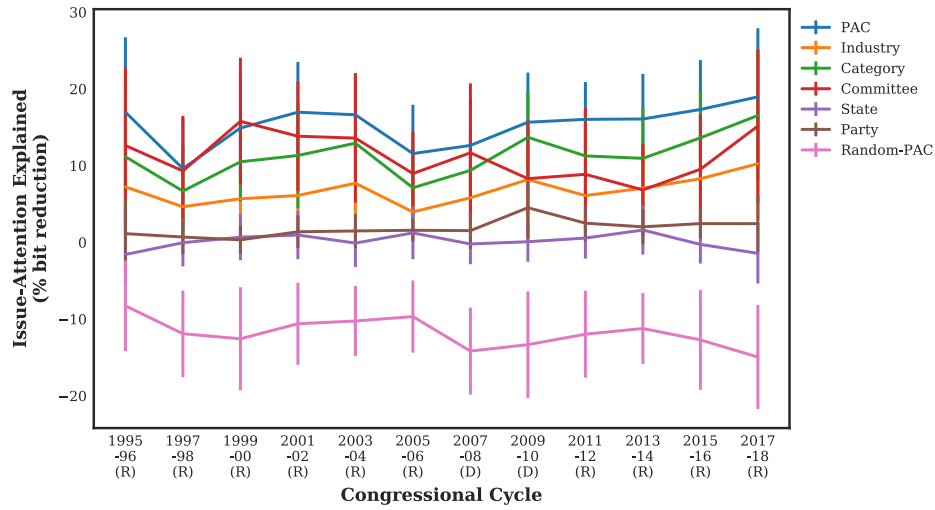


Figure B.3: Held-out set results with error bars when modeling each congressional session separately and predicting legislators’ issue-attention. These correspond to the held-out set results (without error bars) shown in Fig. 3.5b.

the following labels by any one of the two experts are removed: Procedural/Non-Policy, In memoriam, Misc, Regulatory reform, Oversight, Political/non-policy, Regulatory, Constitutionality, Ethics.

The results shown in Fig. B.4 confirm the same pattern that was found using all topics: *PAC* offers the most explanation for legislators’ issue-attention.

Significance testing results. For the model trained on the entire dataset, using the results of the 50-fold cross-validation, we compared the 50 validation set results for the *PAC* attribute with all the other attributes and found that *PAC* exhibits significantly ($p < 0.05$) more predictive value than all the other attributes (compared pairwise) using the Mann-Whitney U statistical test.

As can be seen in both Figs. 3.5a and 3.5b, the top two attributes in terms of their predictive capacity are *PAC*, and *Committee*. For models trained separately on individual Congressional sessions or cycles, we compare each of these to every other attribute using the 30-fold cross-validation results. The comparisons for *PAC*

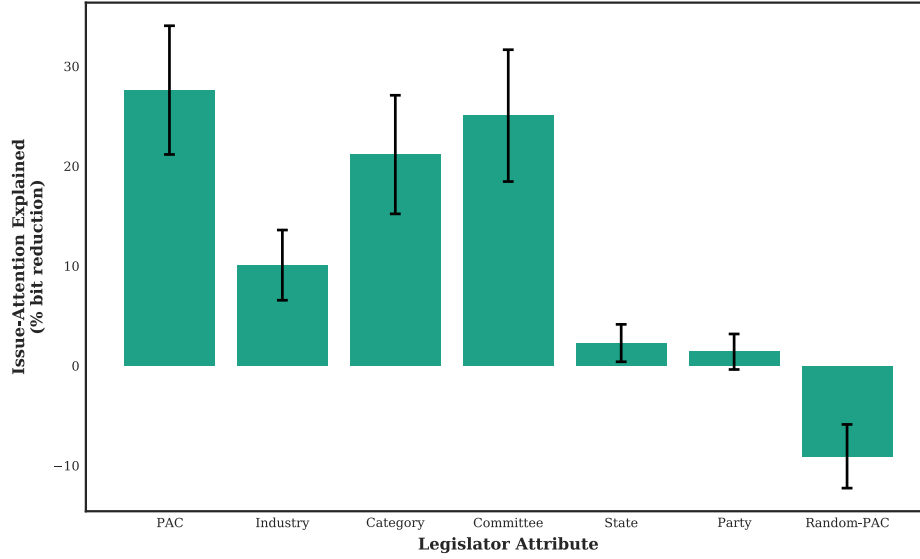


Figure B.4: Comparison for all legislator attributes after removing non-substantial non-policy issues. We find the same pattern as when using all topics: PAC attribute explains legislators’ issue-attention the most.

Congress	PAC >Committee	PAC >Category	PAC >Industry	PAC >State	PAC >Party
1995-96	0.0399	0.0040	0.0000	0.0	0.0
1997-98	0.4036	0.0362	0.0007	0.0	0.0
1999-00	0.4853	0.0018	0.0000	0.0	0.0
2001-02	0.0412	0.0013	0.0000	0.0	0.0
2003-04	0.0560	0.0099	0.0000	0.0	0.0
2005-06	0.0386	0.0026	0.0000	0.0	0.0
2007-08	0.2998	0.0146	0.0001	0.0	0.0
2009-10	0.0001	0.0747	0.0000	0.0	0.0
2011-12	0.0001	0.0013	0.0000	0.0	0.0
2013-14	0.0000	0.0015	0.0000	0.0	0.0
2015-16	0.0000	0.0116	0.0000	0.0	0.0
2017-18	0.0812	0.1259	0.0002	0.0	0.0

Table B.3: Statistical significance test results using Mann-Whitney U test, comparing 30-fold cross-validation results for *PAC* against other legislator attributes. $p < 0.05$ (boldfaced) indicates significantly better results for the PAC attribute against the other attribute for that congressional session.

are shown in Table B.3 and the comparisons for *Committee* are shown in Table B.4; using the Mann-Whitney U statistical test (boldfaced in cases where $p < 0.05$).

Congress	Committee > PAC	Committee > Category	Committee > Industry	Committee > State	Committee > Party
1995-96	0.9614	0.2459	0.0169	0.0000	0.0000
1997-98	0.6021	0.0869	0.0038	0.0000	0.0000
1999-00	0.5206	0.0043	0.0000	0.0000	0.0000
2001-02	0.9601	0.1452	0.0000	0.0000	0.0000
2003-04	0.9457	0.4094	0.0045	0.0000	0.0000
2005-06	0.9626	0.0881	0.0001	0.0000	0.0000
2007-08	0.7053	0.1141	0.0021	0.0000	0.0000
2009-10	0.9999	0.9988	0.4617	0.0000	0.0049
2011-12	0.9999	0.9096	0.0706	0.0000	0.0002
2013-14	1.0000	0.9964	0.5733	0.0001	0.0001
2015-16	1.0000	0.9834	0.2188	0.0000	0.0000
2017-18	0.9210	0.6633	0.0130	0.0000	0.0000

Table B.4: Statistical significance test results using Mann-Whitney U test, comparing 30-fold cross-validation results for *Committee* against other legislator attributes. $p < 0.05$ (boldfaced) indicates significantly better results for the Committee attribute against the other attribute for that congressional session.

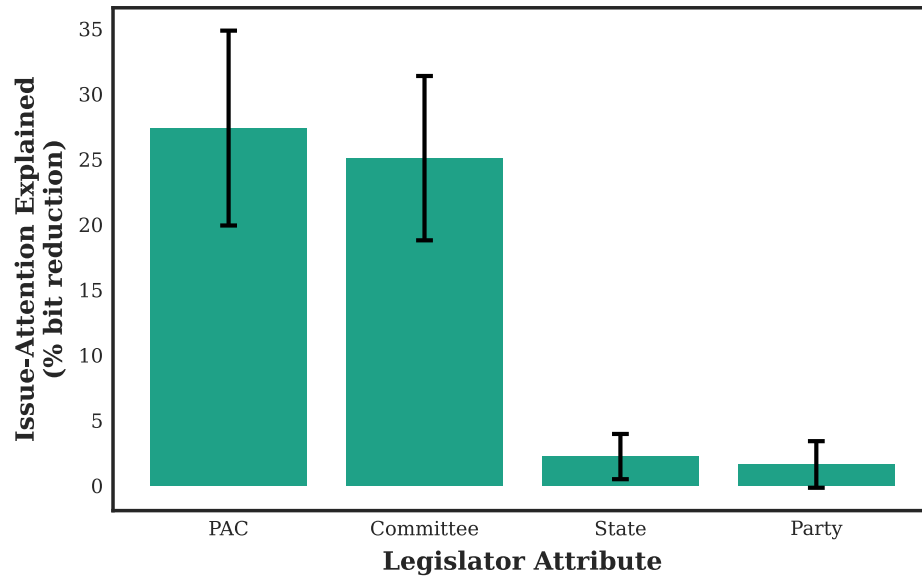


Figure B.5: Held-out set results when issue-attention is quantified using the outputs of an entirely different topic modeling method (as opposed to LDA): the Structured Topic Model (STM).

PAC is significantly more associated with issue-attention, even if issue-attention is quantified using a different topic modeling method. Instead of LDA, we find that reconducting our primary experiments using an entirely different topic model — STM — in order to quantify legislator issue-attention also results in the same finding (Fig. B.5), and our aforementioned significance testing results also continue

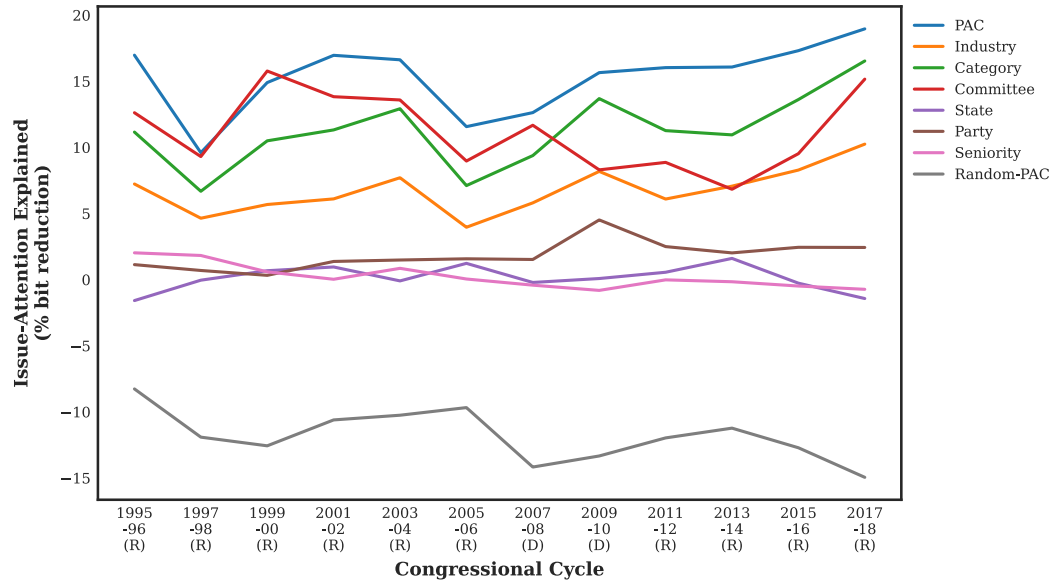


Figure B.6: Held-out set results, congress-by-congress, including results for the additional attribute of legislator *seniority*.

to hold ($N = 50, p < 0.05$).

Legislator seniority does not explain issue-attention. In addition, for experiments considering one congressional session at a time, we assess the potential predictive capacity of legislator *seniority*, following the same one-hot encoding scheme as explained in Section 3.4. Legislator seniority is defined as the number of congressional terms the legislator has served up to and including the particular congressional session being analyzed.

The held-out set results in Fig. B.6 show that much like other attributes like home state and party, seniority does not explain issue-attention, and consistent with other results, our significance testing procedure noted above also finds *seniority* to be significantly worse than the *PAC* attribute ($N = 50, p < 0.05$) across all congressional sessions.

Business PACs are significantly more associated with legislator issue-attention than Labor PACs, but considering them together significantly increases the overall predictive capacity of the PAC attribute. In our dataset, we only consider business and labor PACs. Post-processing, we have 1002 PACs in total (Table 3.1). Out of these, 937 are business PACs and 65 are labor PACs.

Across all legislators in our processed dataset, business PACs donate to at least four legislators, donate to $251.61(\pm 126.73)$ legislators on average, and donate to at most 715 legislators (out of 758).

Across all legislators in our processed dataset, there are 11 labor PACs that do not donate to any legislator. On average, labor PACs donate to $27.67(\pm 19.9)$ legislators, and they donate to at most 61 legislators (out of 758).

Clearly, labor PACs, with their more limited resources as opposed to business PACs, are much more targeted in their donations, which concords with prior research (Gopoian et al., 1984).

We conduct our main logistic regression-based experiment with only business PACs and only labor PACs at a time. Results on the held-out set are provided in Fig. B.7. We find that business PACs offer significantly higher predictive capacity for issue-attention than labor PACs ($N = 50, p < 0.05$). Our significance tests also reveal that considering all PACs explains issue-attention significantly better than considering just business PACs ($N = 50, p < 0.05$).

B.3 Comparing *PAC* and *Committee* Legislator Attributes

After PAC, Committee assignments offer the most predictive value for topic distributions of house representatives. Fig. B.8 offers a closer look at the comparison between the two for the model trained on the entire dataset, and Fig. B.9 provides

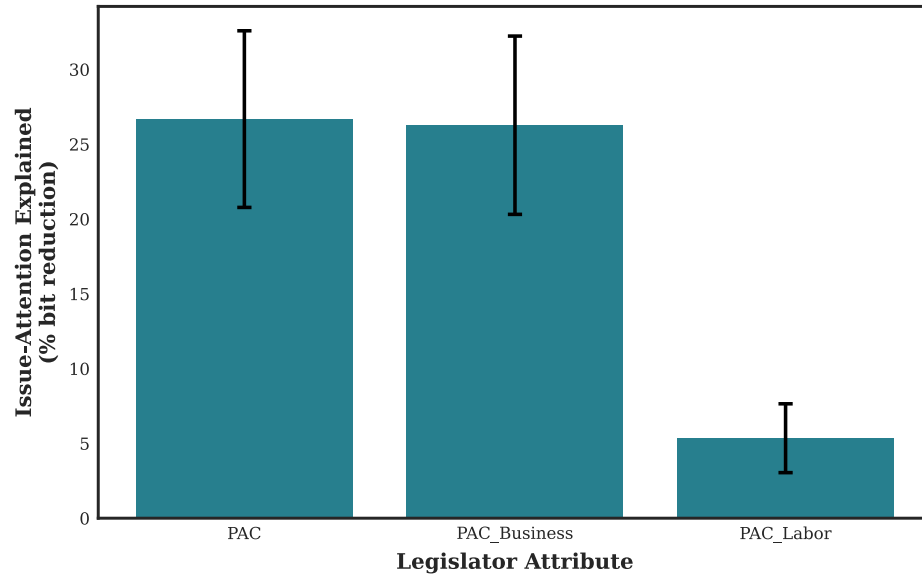


Figure B.7: Comparing predictive capacity for issue-attention when considering business or labor PACs in isolation on the held-out set. Business PACs offer significantly higher predictive capacity than labor PACs. However, our significance test reveals that considering all PACs is still significantly better than considering only business PACs.

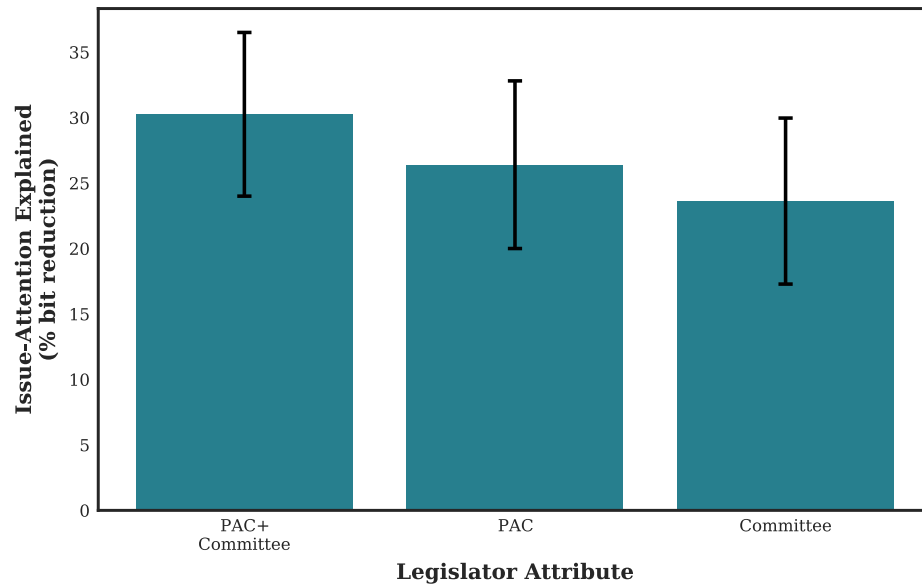


Figure B.8: Specific comparison for the *PAC* and *Committee* legislator attributes (cross-validation results). While *PAC* offers more explanation for issue-attention, results for the combined attribute set of *PAC* and *Committee* suggest some complementary information present in these two explanatory variables.

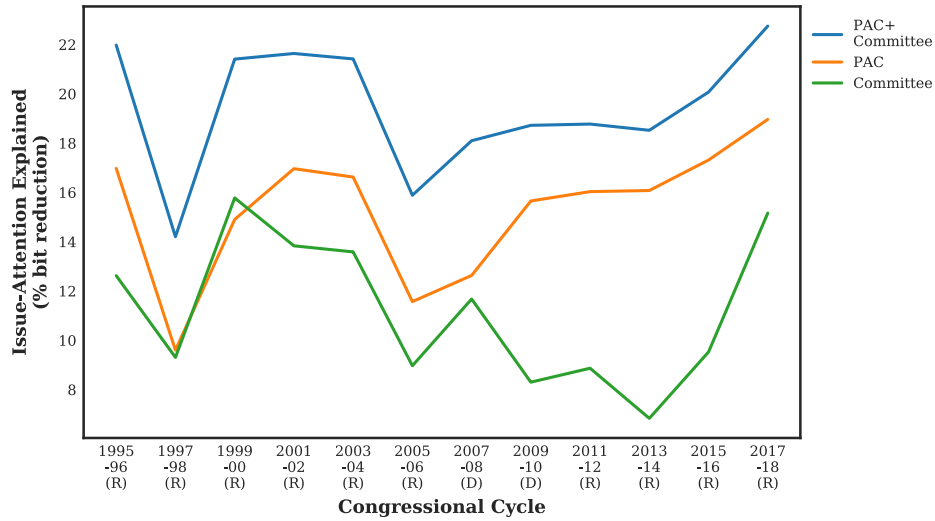


Figure B.9: Specific comparison for the *PAC* and *Committee* legislator attributes — when modeling each congressional session separately (cross-validation results).

the same examination but for models trained separately for each Congressional cycle. We also tried combining all our different feature sets or speaker models by simply concatenating the feature sets to see if the combined models predict topic distributions better. Adding state and/or party information to PACs did not change anything in terms of the predictive capacity, but PAC and Committee attributes, when combined, do show some improvement together (Figs. B.8 and B.9). This indicates there is at least some complementary value in these two sources of information in terms of association with issue-attention in floor speeches.

B.4 PAC-Issue Association Judgment Task Details for Subject-Matter Experts

Annotation task is conducted using a two-step process:

1. Two annotators (political scientists) independently provide labels for each topic based on looking at its most probable words as well as its most probable

Issue (top terms by topic model)	israel, terrorism, terrorist, attacks, attack, terrorists, east, peace, middle, september, middle.east, the.middle.east, threat, terror, syria	education, college, students, programs, training, student, workforce, institutions, opportunities, loans, skills, university, colleges, grants, universities	medical, patients, doctors, hospital, patient, hospitals, doctor, quality, providers, physicians, physician, medicine, centers, emergency, managed
Label (Expert 1)	Foreign Policy (Middle East)	Education (Higher ed)	Healthcare
Label (Expert 2)	Terrorism	Education	Healthcare
(Top PACs) #1	US Beet Sugar Assn	Sears Holdings Corp	Society for Vascular Surgery
#2	Viacom International	Bridgepoint Education	American Psychiatric Assn
#3	Greenberg Traurig LLP	National Education Assn	Cooperative of American Physicians
#4	AECOM US	McDonald's Corp	Society of Thoracic Surgeons
#5	Fluor Corp	TIAA	American Pharmacists Assn
#6	Pizza Hut Franchisees Assn	Prudential Financial	American Assn/Oral & Maxillofacial Surg
#7	American Express	NelNet Inc	American Assn of Neurological Surgeons
#8	American International Group	Navigent Corp	American Academy of Family Physicians
#9	Exelis Inc	American Speech-Language-Hearing Assn	American Society of Plastic Surgeons
#10	American Society of Interventional Pain Physicians	Career Education Colleges & Universities	American Academy of Otolaryngology

Table B.5: Note: Experts (or annotators) created labels for topic using not just the top terms but the top documents (floor speeches) for the topic given by the topic model. And contextualizing info such as sector, industry, category was provided when asking 1-3 associations ratings.

Topic #	Topic Label	PAC Name	Rating	PAC Sector	PAC Industry	PAC Category	PAC Type	
32	Education	Sears Holdings Corp	1	Misc Business	Retail Sales	Department, variety & convenience stores	Business	☐ Given to annotator
32		Bridgepoint Education	3	Other	Education	For-profit Education	Business	
32		National Education Assn	3	Labor	Public Sector Unions	Teachers unions	Labor	☐ Provided by annotator
32		McDonald's Corp	2	Misc Business	Food & Beverage	Restaurants & drinking establishments	Business	
32		TIAA	1	Finance, Insurance & Real Estate	Securities & Investment	Securities, commodities & investment	Business	
32		Prudential Financial	1	Finance, Insurance & Real Estate, Finance, Insurance & Real Estate	Insurance, Insurance	Insurance companies, brokers & agents, Life insurance	Business	
32		NelNet Inc	1	Finance, Insurance & Real Estate	Finance/Credit Companies	Student loan companies	Business	
32		Navigent Corp	1	Finance, Insurance & Real Estate	Finance/Credit Companies	Student loan companies	Business	
32		American Speech-Language-Hearing Assn	3	Health	Health Professionals	Other non-physician health practitioners	Business	
32		Career Education Colleges & Universities	3	Other	Education	For-profit Education	Business	

Figure B.10: Example of how the rating task looks like to an expert annotator - given their own topic label, and 10 PACs, an annotator rated the association of the PAC with the topic on a 1 – 3 Likert scale.

documents (the full protocol and instruction set will be released). Examples of what the topic labels provided by each annotator look like (for the corresponding top words) are shown in Table B.5. Note again that the top documents (based on the document-topic distribution given by the topic model) are also used to come up with the labels.

- Using their own topic label and the top 10 PACs found by the model (examples for certain issues shown in Table B.5), the annotators independently rate

the PACs on a 1 – 3 Likert scale. Sample of the given instructions (more instructions and recommendations/suggestions were provided, the full set will be released with our data, code, and other material):

“For a topic specified by using the topic label assigned in the topic labeling task from before, rate each PAC as being related to the topic (3), or not related to the topic (1) and only in cases where there seems to be some remote possibility of connection so that related or not-related is not clear, assign a (2) rating. PACs are related to a topic if they are a PAC representing an organization concerned with that same issue.

For example, I’d rate the American College of Radiology PAC as related when the topic is Health Care (so a rating of 3), but Trinity Industries PAC (transportation industry) for that same topic would be considered not related (rating of 1).”

Note that the second step was performed for coherent and substantive issue-based topics alone. Topics assigned the following labels in step one were not considered: Procedural/Non-Policy, In memoriam, Misc, Regulatory reform, Oversight, Political/non-policy, Regulatory, Constitutionality, Ethics.

A snapshot of what the final annotation sheet looked like to an annotator rating a topic-PAC association is given in Fig. [B.10](#).

However, in order to use the annotator ratings to judge whether the model is learning something meaningful with respect to the issue-PAC associations captured in the regression weights, we need a baseline. Thus, we created another annotation task in which we randomly assigned 10 (out of the total 1002) PACs to each topic and repeat the two-step process detailed above with the same annotators. The annotators are simply told that these are PAC assignments for the topics from another model (and thus they do not know that these are random assignments a-priori which could have biased the annotations). We used these annotations to

Norm Dicks, the next likely Chairman of the House Defense Appropriations Subcommittee

Norman Dicks (D-WA) has spent his entire career on Capitol Hill, having started out as a Senate staffer for Sen. Warren G. Magnuson, and since 1977, serving as a congressman on the House Appropriations Committee. In March 2010, he was named the Chairman of the Defense Appropriations Subcommittee, replacing the late Rep. John Murtha (D-PA)...

Voting Information

- Mr. Dicks has a liberal Congressional voting record. He voted with a majority of his Democratic colleagues 99.3% of the time during the current Congress. This percentage does not include votes in which Dicks did not vote.

- The differences usually occur on defense spending and foreign policy issues, where he is often more hawkish than the party norm.

On Defense / Military Affairs

- Dicks was a key proponent for the (\$2 billion per plane) B-2 bomber. He successfully ensured money was set aside for the B-2 in the 1995 defense appropriations bill.
- Dicks led select committee investigations into 1998 charges that China had been an illegal recipient of US satellite technology. His report found China to be guilty, with a subsequent report authored by him also alleging that China had stolen significant America nuclear secrets over a 20 year period.
- An early supporter of the Iraq War, by 2005 Dicks u-turned and claimed that he wouldn't have supported the invasion had he not believed Saddam Hussein possessed weapons of mass destruction.
- In 2003, Dicks was a supporter of the sharply increased \$379 billion defense budget – nearly \$50 billion higher than that of 2002. Despite the large increase, Dicks asserted that yet more money was needed urgently to modernize US military hardware.

Figure B.11: An example of the kind of real-world relationships that the outputs of our modeling and analysis can capture automatically - for a PAC (**BAE Systems**) and an issue (**International Security/Foreign Policy**) that our model (as well as experts) deem related to one another, we find that in a particular congressional cycle (1995-96), there is a significant temporal connection between this PAC and the issue, i.e, this PAC donates a significant amount closer to speeches on the particular issue of international security and foreign policy in 2015-16 than for speeches that are not on this issue. In particular, the legislator - **Norm Dicks (D-WA)** - emerged as a frequent recipient of such temporally significant donations by this PAC, and a simple search on a search engine reveals that this connection is backed by what is known to journalists and validated by real-world knowledge in a straightforward manner. The snapshot in the image is taken from <https://armscontrolcenter.org/norm-dicks-the-next-likely-chairman-of-the-house-defense-appropriations-subcommittee/>.

compare and establish that the model is indeed finding meaningful associations, and the way the learned weights rank the PACs for a topic is more meaningful to domain experts than if the PACs were chosen randomly (Section 3.6).

Per both the Mann-Whitney U test and the Wilcoxon signed-rank test for comparing expert ratings for top 10 PACs per our model versus randomly selected 10 PACs, for both experts, the model-selected top PACs are rated significantly higher than the randomly-selected PACs ($p < 0.00001$). All of this clearly shows that the regression model is learning meaningful topic-PAC associations, and apart from better quantitative predictive performance as compared to baselines such as having the party or state information, our model can also help get which PACs are interested in targeting which topics.

PAC name	Topic label (annotator 1/ annotator 2)	Speech text snippet	Legislator (Affiliation), Speech Date	Donation to legislator (Amount, Date)	Mean (std. dev.) donation amount for the donor in the election cycle
General Dynamics	Defense/ Military	Mr. Speaker, I thank the gentleman for yielding me time. Mr. Speaker, I just wanted to state, so my colleagues will understand very clearly, and I think the gentleman from Pennsylvania [Mr. Weldon] made the point, that this President does not want to defend the United States against incoming ballistic missiles. That was his major objection to this bill, along with the idea that he also wants to have the right to delegate to foreign commanders the command of U.S. troops. We are now going to enter a period in which it is important for Members of this House who feel that defense is important to enter a full-court press this year to develop defenses against incoming ballistic missiles, both for the people of the United States and for our troops in theater. We are going to do this. The President has given up his most solemn responsibility, and that is to defend the people of the United States of America, and he is denied that responsibility in this bill.	Duncan L. Hunter (R-CA) on 1996-01-24	2000 USD, 1996-01-21	720 USD (±470 USD)
AFLAC Inc (Accident and health supplemental insurance company;)	Economic/ Economy (jobs)	Mr. Speaker, as a result of our arcane and complicated pension laws, 70 million workers have no pension plan. Unfortunately, Americans who work in small businesses are much less likely to have pension coverage than those who work for larger companies. Among companies with fewer than 100 employees, as many as 80% of the workforce have no retirement savings plan available to them. The primary cause: small business owners find the cost and complexity of setting up and maintaining retirement plans to be overwhelming. So last year, Congress passed the Portman-Cardin pension reforms to help workers save for their future and enable small businesses to offer pension plans to their employees. The changes we made streamline and simplify the complex rules governing our pension system to ensure meaningful coverage of small business employees. They will reduce the administrative burden on small businesses and provide incentives to help them establish plans for their workers, including cutting the IRS user fee small businesses have to pay to establish a pension plan and lowering premiums small businesses pay for their defined benefit plans to make that option more attractive ...	Nancy Johnson (R - CT) on 2002-06-21	5000 USD, 2002-06-14	1690 USD (±1220 USD)
BAE Systems	International Security/ Foreign Policy	Mr. Chairman, I thank the gentleman from Wisconsin for yielding time to me. First let me say that I agree very much, this is an American, this is a NATO conflict. We in this House should speak with one voice and not be putting it on political terms. I feel very, very deeply about this. I support thisbill. At the end of the day, I support this bill. It is a major step toward my goal of making this the year of the troops, the year in which we recognize the needs of those who serve in uniform. I also support it because it ensures that our military has more than adequate resources to carry out the Kosovo air campaign. It bolsters the military readiness of our forces in the Balkan theater and the Armed Forces as a whole. It provides the resources to help address the tragic humanitarian situation in Kosovo. The basis of this bill was a \$6 billion administration request in emergency funding. The request was based on four categories, military operations in and around Kosovo, Kosovar refugee relief, munitions and readiness munitions, and Desert Thunder and Desert Fox military operations. In addition to the administration's original request, our colleagues on the Committee on Appropriations have seen fit to add to the President's request, both to the humanitarian request and the matter request. There are some problems that our colleagues had on the Committee on Appropriations, and they have tried to address them. They have added certain categories. Mr. Chairman, allow me to comment on two major additions to the original request. First, this bill sends the right signal to our men and women in uniform by providing \$1.8 billion to fund the administration's military pay and retirement package, of course, conditioned upon the enactment of authorizing legislation through our Committee on Armed Services. Second, this bill provides for \$1.1 billion in unrequested funds for overseas military construction in Europe and Southeast Asia. The inclusion of these projects is similar to the inclusion of the administration's pay and retirement package.	Ike Skelton (D - MO) on 1999-05-06	3000 USD, 1999-05-07	890 USD (±620 USD)

Table B.6: After selecting relevant PAC-issue pairs with potential temporal connection for the cycle: some examples of speeches made on the topic and cases of the relevant PAC making a donation much higher than their mean donation amount during that congressional cycle to the legislator giving that speech within a particular time window around the speech.

B.5 More Examples and Results of Discovered Temporal Connections

Examples of significantly connected speech and donation events occurring in temporal proximity are shown in Table B.6.

By ranking and scoring these events, our procedure (Section 3.7) can automati-

Big questions for CUs as Royce retires from House

By **Aaron Passman** January 10, 2018, 1:20 p.m. EST 6 Min Read



During a Congressional cycle in which a host of long-serving legislators have [announced their](#)



[intentions to retire](#) or step down, one exit may sting the credit union movement more than the



rest.



U.S. Rep. Ed Royce (R-CA) announced Monday his intention to [retire from Congress](#) at the end of this term, who has earned a reputation for being a credit union champion, having introduced a number of pro-credit union bills during his tenure.

Figure B.12: An example of the kind of real-world relationships that the outputs of our modeling and analysis can capture automatically - for a PAC (**National Assn of Federally Insured Credit Unions**) and an issue (**Finance/Financial regulation**) that our model (as well as experts) deem related to one another, we find that in two different congressional cycles (2009-10 and 2015-16), there is a significant temporal connection between this PAC and the issue, i.e, this PAC donates a significant amount closer to speeches on the particular issue of education in 2015-16 than for speeches that are not on this issue. In particular, the legislator - **Ed Royce (R - CA)** - emerged as the recipient of such temporally significant donations by this PAC in both cycles, and a simple search on a search engine reveals that this connection is backed by what is known to journalists and validated by real-world knowledge in a straightforward manner. Snapshot taken from <https://www.americanbanker.com/creditunions/news/big-questions-for-credit-unions-as-royce-retires-from-house>.

cally uncover some real-world donor-legislator relationships which are easily validated by straightforward search engine queries; illustrative examples are provided in Figs. [B.11](#) and [B.12](#).

Appendix C: Context-based Variation in Communicative Choices: US Legislators’ Ideological Expression Across Audi- ences

C.1 Preprocessing Details

For the most part, we followed the text processing settings described in Appendix B in [Vafa et al. \(2020\)](#), and some major differences are already described in “Preparing Text Data” in Section 4.4.

In the case of floor speeches, speakers giving less than 25 speeches were removed, and terms used by less than 50 representatives in the data were also removed when processing and finalizing the vocabulary.

We also remove procedural speeches - short speech segments containing pro forma language alone such as “I yield back the balance of my time”, following our own implementation of the procedure to identify such speeches laid out in [Card et al. \(2022\)](#). Their procedure works with the assumption that “...short speeches which are repeated verbatim or nearly verbatim many times should be treated as procedural” — after some manual and heuristic based identification of a few positive and negative samples, a regularized logistic regression classifier was able to identify such procedural speeches with about 99% accuracy. We refer the reader to [Card et al. \(2022\)](#) for details (“identifying procedural speeches” in

their supplementary information) as well as our publicly available implementation. Around 22k speech texts out of an initial $\sim 81k$ speeches were removed as procedural speeches due to this step.

The speeches containing only pro-forma language can be misleading and negatively affect ideal point modeling since they would not demonstrate partisan language. We subjectively find that removing these speeches improves the partisan topics discovered and recommend this step for application of TBIP in future work.

In case of tweets, terms used by less than 5 representatives in the data are also removed when processing and finalizing the vocabulary. In both datasets, unigrams, bigrams, and trigrams are allowed.

Our final processed dataset for floor speeches contains 58,932 speeches and 11,433 unique terms in the vocabulary. For Twitter data, our final processed dataset contains 295,327 tweets and 9,343 unique terms in the vocabulary. For both datasets, we run the TBIP model with 50 topics.

The new dataset we put together to enable this work including model outputs, human annotation instructions and results to establish model viability, and our entire codebase for applying the model and obtaining the outputs at the heart of our analyses is publicly available at <https://zenodo.org/record/7641732> (DOI: 10.5281/zenodo.7641732).

C.2 Stability of Ideal Point Estimates

Given a set of initial topics (estimated in the first step) and for a given dataset, we assess the stability of the ideal point values for the authors (legislators) estimated by TBIP by running the model 10 times, and then for all pairs of runs, we look at the correlation between the ideal point estimates of the authors. This is done because

some variables in the estimation process are randomly initialized and it is important to know how much of a variance it can cause in the final TBIP estimates the model outputs. This allows us to assess if one can reliably use the model results of a single run.

The ideal point estimates learned on multiple runs (with different random seed) are highly correlated across pairs of runs (average Spearman $r = 0.675$, $p < 0.0005$ and average Pearson $r = 0.713$, $p < 0.0005$ for floor speeches; average Spearman $r = 0.897$, $p < 0.0005$ and average Pearson $r = 0.908$, $p < 0.0005$ for tweets).

Note that we did the above stability analysis with the original Poisson estimation-based topic initialization, but we expect the stability results to be unaffected by initializing with MALLET LDA as described in Section 4.5 since ideal point estimates correlate highly (average Pearson $r > 0.9$) when comparing Poisson vs MALLET initialization (though topics discovered are subjectively determined to be more coherent).

The TBIP model, therefore, appears to be quite stable in terms of ideal point values it discovers for authors.

C.3 Details of Expert Assessment Tasks to Validate Automated Outputs

Apart from stability, we also systemically establish validity of the TBIP model outputs in order to ensure the model can be reliably used for empirical political science research.

Four subject-matter experts (students pursuing their doctoral degree in political science) separately analyzed the coherence and polarization of the *tbip* topics — both the topics provided by MALLET used for initialization and the learned ideological

topics derived by the TBIP estimation process. Below is an overview of the directions given to the annotators and the results of their analysis.

C.3.1 Initial Topic Assessment (Pre-TBIP)

Coherence Rating	Floor Speeches		Tweets	
	Annotator 1	Annotator 2	Annotator 1	Annotator 2
1.0	4%	16%	0%	22%
2.0	2%	6%	2%	8%
3.0	94%	78%	98%	70%
Cohen's Kappa inter-annotator agreement	0.151		0.040	

Table C.1: Initial topics used to initialize TBIP, rated for coherence by two subject-matter experts.

Polarization Rating	Floor Speeches		Tweets	
	Annotator 1	Annotator 2	Annotator 1	Annotator 2
1.0	44%	46%	48%	52%
2.0	0%	8%	0%	10%
3.0	52%	30%	52%	16%
N/A	4%	16%	0%	22%
Cohen's Kappa inter-annotator agreement	0.370		0.221	

Table C.2: Initial topics used to initialize TBIP rated at least 2 on the 3-point Likert scale for coherence, then rated for expected ideological polarization by two subject-matter experts.

Coherence: Rate the topic, on a 1-3 scale, on its coherence (1 being not coherent, 3 being very coherent). Does that topic represent an easily identifiable category or a meaningful concept? The top words for the topic, along with the top documents associated with that topic help make this judgment. Broadly, a set of items can be said to be coherent if they enable human recognition of an identifiable category when viewed together. The results are shown in Table C.1.

Expected Polarization: If the topic is rated above 1 for coherence, rate the topic, on a 1-3 scale (1 being not polarized, 3 being polarized), on the expected polarization of this topic: do you expect meaningful ideological differences in the way liberals and conservatives would talk about the category or concept or issue or the stance they would hold on that issue? The top words for the topic, along with the top documents

Perspective a)	Floor Speeches		Tweets	
	Annotator 1	Annotator 2	Annotator 1	Annotator 2
IS about <Issue Label>	90.48%	83.33%	92.5%	57.5%
MIGHT be about <Issue Label>	7.14%	14.29%	7.5%	40.0%
IS NOT about <Issue Label>	2.38%	2.38%	0.0%	2.5%
Cohen's Kappa inter-annotator agreement	0.806		0.594	

Table C.3: Rating if the provided topic label (created using initial topics) applies to the the binary ideological perspective (a) for that same topic post-TBIP.

Perspective b)	Floor Speeches		Tweets	
	Annotator 1	Annotator 2	Annotator 1	Annotator 2
IS about <Issue Label>	90.48%	80.95%	75.0%	65.0%
MIGHT be about <Issue Label>	9.52%	14.29%	22.5%	30.0%
IS NOT about <Issue Label>	0.0%	4.76%	2.5%	5.0%
Cohen's Kappa inter-annotator agreement	0.830		0.619	

Table C.4: Rating if the provided topic label (created using initial topics) applies to the the binary ideological perspective (b) for that same topic post-TBIP.

associated with that topic, as well as your personal knowledge of American politics will help make this judgment. The results are shown in Table C.2.

Based on the topic labels after consensus between the two annotators (in the pre-TBIP stage), we filtered topics that were not coherent or represented non-substantive issues, which left us with 42 topics for speeches and 40 topics for tweets that were then presented to two different annotators in the post-TBIP annotation stage.

C.3.2 Post-TBIP Output Assessment — Ideological Topics

Label Applicability to Perspectives: Looking at the top words shown for the perspective a) of a particular issue number (columns A and B under the issue number), select one of the three options based on your judgment of whether the Issue Label is

a) and b) represent polarized perspectives for the issue?	Floor Speeches		Tweets	
	Annotator 1	Annotator 2	Annotator 1	Annotator 2
YES	31.71%	35.9%	28.21%	42.11%
SOMEWHAT	41.46%	30.77%	28.21%	31.58%
NO	24.39%	17.95%	43.59%	15.79%
UNSURE	2.44%	15.38%	0.0%	10.53%
Cohen's Kappa inter-annotator agreement	0.330		0.435	

Table C.5: Rating if the two ideological perspectives on an issue derived using TBIP represent polarized or partisan framings of the issue.

a suitable label for the underlying concept or category you think the words shown for 'a)' represent. Put differently, is the shown <Issue Label> applicable to the words shown for the words under 'a)' for that issue? Results are shown in Table C.3.

The same question was asked for perspective *b*). Results are shown in Table C.4.

Ideological Polarization: Looking at both the perspectives *a*) and *b*), the experts were asked if they represent ideologically polarized perspectives. Results are shown in Table C.5.

C.3.3 Summary of results for floor speeches

For floor speeches, 81.82% of initial topics (before any ideal point modeling) expected to have political polarization by either annotator were deemed at least somewhat ideologically polarized when looking at the polarized perspectives (provided by TBIP) by either annotator.

For the issues that the annotator deemed ideologically polarized when looking at the two perspectives provided by TBIP:

- For issue perspective that were the liberal perspective per TBIP, annotator 1 identified 90.6% of them as liberal, and they were unsure about the remaining 9.4%.
- For issue perspective that were the conservative perspective per TBIP, annotator 1 identified 78.1% of them as conservative, and they were unsure about the remaining 21.9%.
- For issue perspective that were the liberal perspective per TBIP, annotator 2 identified 67.7% of them as liberal, 8.8% as conservative, and they were unsure about the remaining 24%.

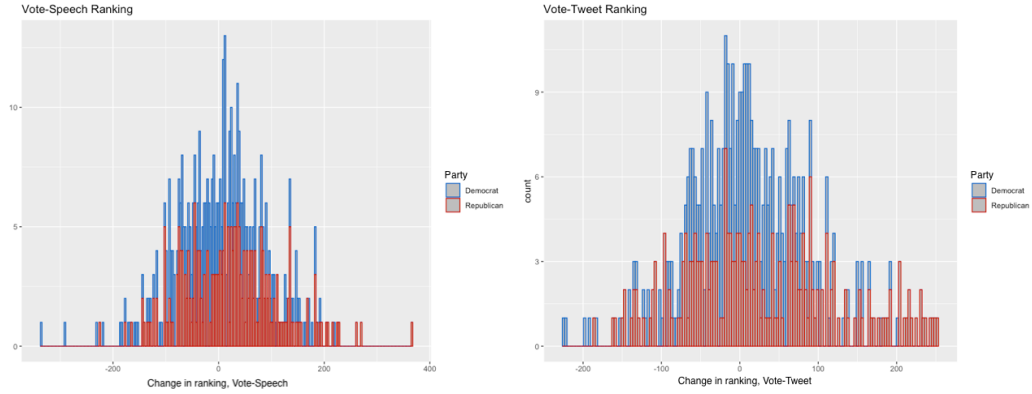
- For issue perspective that were the conservative perspective per TBIP, annotator 2 identified 67.7% of them as conservative, 8.8% as liberal, and they were unsure about the remaining 24%.

C.3.4 Summary of results for tweets

For tweets, 72.73% of initial topics (before any ideal point modeling) expected to have political polarization by either annotator were deemed at least somewhat ideologically polarized when looking at the polarized perspectives (provided by TBIP) by either annotator.

For the issues that the annotator deemed ideologically polarized when looking at the two perspectives provided by TBIP:

- For issue perspective that were the liberal perspective per TBIP, annotator 1 identified 96.7% of them as liberal, and they were unsure about the remaining 3.3%.
- For issue perspective that were the conservative perspective per TBIP, annotator 1 identified 76.7% of them as conservative, 6.7% as liberal, and they were unsure about the remaining 16.6%.
- For issue perspective that were the liberal perspective per TBIP, annotator 2 identified 84.9% of them as liberal, and they were unsure about the remaining 15.1%.
- For issue perspective that were the conservative perspective per TBIP, annotator 2 identified 84.9% of them as conservative, and they were unsure about the remaining 15.1%.



(a) Independent-level shift from vote-based ideal points to floor-speech-based ideal points.
 (b) Independent-level shift from vote-based ideal points to tweet-based ideal points.

Figure C.1: Independent-level shift in member rankings from vote-based to text-based (floor speeches and tweets) ideal points, 115th and 116th Congress.

C.4 Extended Results

Results with the full congressional ordinary least squared (OLS) regression with interactive variable for the party are shown in Table C.6.

Changes in the ranked ideal point based position for members across venues with the interactive OLS regression model are shown in Table C.7.

The results for independent-level shift from vote-based to speech-based ideal points for members are shown in Fig. C.1a, and from vote-based to tweet-based ideal points in Fig. C.1b.

<i>Independent Variable</i>	<i>Dependent variable:</i>		
	Vote	Speech	Tweet
Constant	-1.171*** (0.128)	-0.657 (0.613)	-2.030*** (0.353)
House Candidate Vote Share, 2016	-0.003*** (0.001)	0.004 (0.004)	0.006* (0.003)
House Candidate Vote Share, 2018	0.001 (0.001)	-0.002 (0.005)	-0.0002 (0.003)
Party, Republican	1.829*** (0.179)	2.193** (0.857)	2.376*** (0.495)
Trump Vote Share, 2016	0.005*** (0.001)	0.008 (0.007)	0.024*** (0.004)
District percent white	0.0002 (0.001)	-0.001 (0.003)	-0.001 (0.001)
District unemployment rate	0.011 (0.011)	0.009 (0.054)	0.018 (0.031)
Urban	0.048 (0.032)	-0.120 (0.152)	-0.083 (0.088)
Suburban	0.007 (0.026)	-0.217* (0.124)	-0.066 (0.071)
Number of House Terms	-0.003 (0.001)	-0.041*** (0.008)	0.023*** (0.004)
MC Gender	-0.011 (0.018)	-0.075 (0.089)	-0.115** (0.052)
Problem Solvers Caucus	0.007 (0.037)	0.071 (0.071)	-0.048 (0.096)
Progressive Caucus	-0.026 (0.020)	-0.161* (0.095)	-0.145*** (0.055)
New Democratic Coalition	0.035* (0.021)	0.050 (0.099)	-0.004 (0.084)
Blue Dog Coalition	0.294*** (0.031)	0.561*** (0.147)	0.568*** (0.057)
RSC	0.139*** (0.022)	0.188* (0.106)	0.148** (0.062)
Freedom Caucus	0.069*** (0.023)	-0.141 (0.112)	-0.025 (0.065)
Democratic Party leader	-0.042 (0.028)	-0.218 (0.133)	-0.148* (0.076)
GOP Party leader	0.005 (0.036)	-0.325* (0.172)	-0.027 (0.099)
Member of top committee	-0.006 (0.028)	-0.404*** (0.103)	-0.081 (0.059)
Committee Chair	0.038 (0.035)	0.281* (0.168)	-0.232** (0.097)
House Candidate Vote Share, 2016*Republican	0.004** (0.002)	-0.002 (0.008)	-0.001 (0.004)
House Candidate Vote Share, 2018*Republican	0.001 (0.002)	-0.004 (0.011)	-0.004 (0.006)
Trump Vote Share, 2016*Republican	-0.002 (0.001)	-0.006 (0.011)	-0.015** (0.003)
District percent white*Republican	-0.003** (0.002)	-0.006 (0.005)	-0.003 (0.003)
District unemployment rate*Republican	0.011 (0.019)	0.057 (0.095)	0.028 (0.055)
Urban district*Republican	-0.006 (0.117)	-0.381 (0.560)	-1.446*** (0.321)
Suburban District*Republican	0.006 (0.006)	-0.013 (0.162)	-0.011 (0.094)
MC Gender*Republican	0.002 (0.034)	-0.076 (0.185)	0.138 (0.107)
Problem Solvers*Republican	-0.059 (0.039)	-0.255 (0.184)	0.007 (0.109)
Top Committee*Republican	-0.005 (0.033)	-0.353** (0.155)	0.040 (0.089)
Committee Chair*Republican	-0.036 (0.053)	-0.325 (0.025)	0.325** (0.145)
R ²	0.991	0.768	0.930

Table C.6: **Full Congressional OLS model, 115th – 116th Congress.** Party (Republican) acts as an interaction variable for those relevant variables (any independent variable that applies to both Democratic and Republican members). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$; MC = Member of Congress.

<i>Independent Variable</i>	<i>Dependent Variable</i>	
	Shift Vote to Speech	Shift Vote to Tweet
Republican MC	92.184 (100.61)	191.165** (92.342)
Female MC	18.709 (11.673)	15.573* (8.848)
Republican * Female MC	-30.956 (19.699)	-5.249 (22.111)
Non-White MC	6.379 (14.544)	17.029* (10.428)
Republican * Non-White MC	-18.972 (26.659)	42.126 (34.843)
Number of terms served	-3.062** (1.529)	-3.739*** (1.365)
Republican * Number of terms served	6.609*** (1.989)	-2.686 (2.017)
Percent vote in last election	0.337 (0.534)	-.433 (0.503)
Republican * Percent vote last election	0.936 (.848)	1.840* (1.054)
District vote for Trump	0.452 (0.970)	-0.396 (0.761)
Republican * District vote for Trump	-0.140 (1.344)	1.034 (1.336)
Suburban district	50.364*** (16.577)	25.325* (14.534)
Republican * Suburban district	-18.444 (21.248)	-18.454 (19.591)
Urban district	32.234 (21.329)	40.091** (17.173)
Republican * Urban district	17.271 (31.969)	41.765 (61.644)
District percent white	0.336 (0.455)	0.095 (0.395)
Republican * District percent white	-1.367* (0.716)	-2.225*** (0.703)
District unemployment rate	-12.128 (7.510)	-6.720 (6.772)
Republican * District unemployment	15.138* (9.013)	-29.687*** (8.992)
Problem Solvers Caucus	-28.163** (14.002)	-25.213*** (10.185)
Republican * Problem Solvers Caucus	12.493 (21.300)	-7.128 (18.965)
Republican Study Group (Rep)	24.492** (11.885)	9.782 (13.960)
Freedom Caucus (Rep)	73.036*** (15.839)	29.549* (16.720)
Progressive Caucus (Dem)	21.389* (12.584)	14.242 (11.233)
New Democratic Coalition (Dem)	7.381 (14.164)	13.615 (11.511)
Blue Dog Coalition (Dem)	4.293 (18.828)	7.683 (9.534)
Party leader	11.264 (18.253)	10.630 (17.528)
Republican * Party leader	25.750 (19.332)	-27.521* (14.772)
Committee chair	-40.698* (24.261)	0.741 (23.652)
Republican * Committee chair	55.491* (32.748)	-26.406 (28.826)
Top committee member	10.535 (10.580)	5.448 (8.627)
Republican * Top committee	-7.748 (16.206)	-50.197*** (14.648)
Constant	-58.703 (73.762)	30.130 (62.414)
Observations	489	471
R ²	0.268	0.349

Table C.7: **Changes in Ranked Ideal Point Position Interactive OLS model, 115th – 116th Congress.** *Negative sign indicates more conservative than vote, positive sign indicates more conservative than vote.* * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$; MC = Member of Congress.

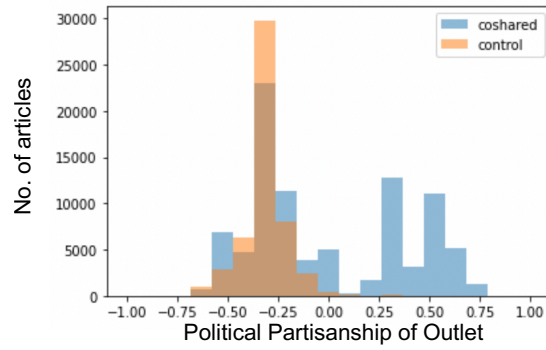
Appendix D: Consequences of Communicative Choices: Mainstream News Articles Co-Shared with Fake News Buttress Misinformation Narratives

D.1 Distribution of Article Counts per Outlet in Co-shared & Control Group

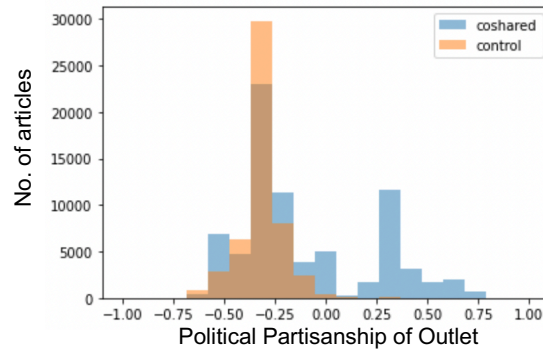
When considering all reliable outlets (Fig. [D.1a](#)) and only trustworthy outlets (Fig. [D.1b](#)), right-leaning outlets are over-represented in the co-shared group (in terms of the number of articles), even when both groups of articles (co-shared and control) always consist of articles published by the same set of outlets. This ideological skew is not present in our third condition of only considering liberal outlets (Fig. [D.1c](#)). Political scores for outlets use annotations created in [Green et al. \(2021\)](#), as described in Section [5.2](#) (see criteria number three for mainstream media).

D.2 Partisanship Explanation Analysis: Procedure and Results

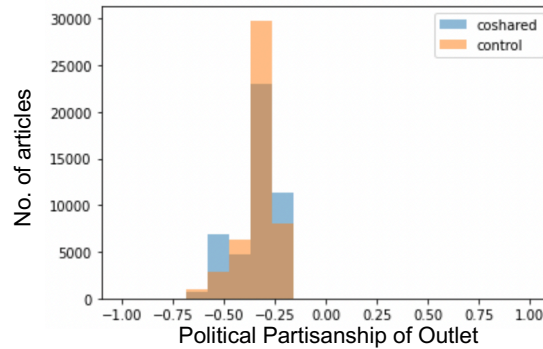
Based on article-level ideological or partisanship scores derived from the relative number of Democrats and Republicans sharing the article in the Twitter panel dataset (where voter registration data is linked with the user tweets; see [Hughes et al. \(2021\)](#); [Shugars et al. \(2021\)](#) for details), we find that the co-shared set of mainstream articles is significantly more conservative (or significantly more popular



a) All reliable outlets



b) Trustworthy outlets



c) Liberal outlets

Figure D.1: We plot the number of articles in co-shared and control groups against the political partisanship score of the outlet (derived from party affiliation of the Twitter audience sharing political content from the outlets by prior work in [Green et al. \(2021\)](#)).

among Republicans) compared with the control group of mainstream articles (independent t -test, $p < < < 0.05$). Partisan curation thus presents itself as another viable explanation, and we examine the effect of misinformation narratives in the presence of partisan curation as a competing, but potentially complementary, explanation in

Type of outlets	Misinformation narrative library	Dimensionality of extracted narratives	n	
All reliable outlets	All fake news narratives	Low-dimensional	91219	$p > 0.05$
		High-dimensional	64090	$p > 0.05$
	Recurring fake news narratives	Low-dimensional	1537	$p << 0.05$
		High-dimensional	580	$p > 0.05$
	False-claims narratives	Low-dimensional	2599	$p <<< 0.05$
		High-dimensional	165	$p > 0.05$
Only trustworthy outlets	All fake news narratives	Low-dimensional	48417	$p <<< 0.05$
		High-dimensional	9373	$p <<< 0.05$
	Recurring fake news narratives	Low-dimensional	1388	$p <<< 0.05$
		High-dimensional	445	$p <<< 0.05$
	False-claims narratives	Low-dimensional	2142	$p <<< 0.05$
		High-dimensional	138	$p > 0.05$

Table D.1: Results of significance testing showing that across various categories of reliable mainstream news outlets when using our graph-based method and entire method (Section 5.3) starting with the subset of all Twitter data where tweet authors are conservatives (registered Republicans) only.

various ways outlined below.

Repeating our main experiment with a Republicans- or conservatives-only Twitter panel. We create a subset of our Twitter data to only contain tweets by registered Republicans, and then repeat our entire experiment – creating the URL-to-URL graph, finding co-shared and control sets of articles, extracting narratives from the combined set of fake news, false claims, co-shared and control mainstream groups of texts, and comparing the presence of narratives derived from misinformation content in co-shared versus control texts.

The results are presented in Table D.1. We find that while the number of cases where there is a significant difference between the two groups of mainstream articles is reduced, it nevertheless remains a significant factor in many cases, especially when considering mainstream articles published by trustworthy outlets.

We, therefore, compare the two explanations in a regression setup to see if

	coef	std err	z	P >—z—
(Intercept)	1.7467	0.015	118.895	$< 2e - 16^{***}$
$X_{\text{partisanship}}$	3.0532	0.022	141.193	$< 2e - 16^{***}$
$X_{\text{misinfo_narrative}}$	0.1791	0.006	31.729	$< 2e - 16^{***}$

Table D.2: Logistic regression results for $y \sim X_{\text{partisanship}} + X_{\text{misinfo_narrative}}$, where $y = 1$ when the mainstream article is co-shared, and $y = 0$ when the mainstream article is control. A higher partisanship score means more conservative. The results show that an article being more conservative (measured by relative popularity among Republicans and Democrats in terms of their sharing behavior on Twitter) and an article having a higher number of misinformation narratives are significantly and positively associated with co-sharing likelihood. *** indicates significance at $p < 0.001$.

misinformation narratives constitute a significant factor in distinguishing co-shared and control groups (for articles shared in our entire Twitter data), as well as what value addition (in terms of model-fit) might misinformation narratives provide over-and-above the partisan curation, *i.e.*, is partisanship alone enough and subsumes the narrative-based explanations, or are the two distinct, complementary factors?

Regression experiment using both partisanship and narrative-based factors.

With every instance being a mainstream article, we set up a logistic regression expression where the dependent variable (y) is a binary indicator of being a co-shared or control article. In our main logistic regression setup, one independent variable (X_1 or $X_{\text{partisanship}}$) is the partisanship score of the article (the relative sharing by Republicans and Democrats as mentioned above). We combine all recurring narratives in fake news and all narratives in false fact-checked claims as our misinformation narrative set, and then count the number of narratives in the mainstream article that are also present in the misinformation narrative set. This count forms the second independent variable (X_2 or $X_{\text{misinfo_narrative}}$).

Partisanship being a significant explanation is something we expect, but our results in Table D.2 show that misinformation narratives offer another significant

	K	AICc	Delta_AICc	AICcWt
full model (partisanship + misinformation narratives)	3	128053.1	0.00	1
partisanship-only model	2	129156.7	1103.64	0

Table D.3: We run two logistic regression models, the full model where both factors are used as independent variables ($y \sim X_{\text{partisanship}} + X_{\text{misinfo_narrative}}$), and partisanship-only model ($y \sim X_{\text{partisanship}}$); where $y = 1$ when the mainstream article is co-shared, and $y = 0$ when the mainstream article is control. We compare the model fit using the information-theoretic concept of Akaike information criterion (AIC), implemented using the R library ‘AICcmodavg’ (Mazerolle, 2023). In the table, K is the number of parameters of the model, AICc is the information score of the model (lower indicates better fit), Delta_AICc is the difference in the AIC score between the best model and the model being compared, and AICcWt is the proportion of the total amount of predictive power provided by the full set of models contained in the model being assessed (see: <https://www.scribbr.com/statistics/akaike-information-criterion/> for more details).

explanation. We also check if these two independent variables are co-linear, by conducting a multicollinearity test based on the variance inflation factor (VIF) test (Marquardt, 1970; Akinwande et al., 2015). We find that for these two variables, the VIF is 1.013826, which indicates that these two factors are not co-linear or only slightly co-linear (Akinwande et al., 2015). The Spearman correlation value is also low ($r = 0.1$).

All of this suggests that partisanship and misinformation narrative presence in mainstream articles are two significant and complementary factors for distinguishing between co-shared and control mainstream articles, and partisanship as a factor does not explain away misinformation narratives as an explanatory factor.

Using information theory to quantify the explanation value provided by the misinformation narratives factor. Finally, we investigate whether having misinformation narratives as a factor offers significant additional explanatory value (towards distinguishing co-shared and control mainstream articles) above and beyond what partisanship already offers. Apart from the full regression model

($y \sim X_{\text{partisanship}} + X_{\text{misinfo_narrative}}$), we also run a partisanship-only model ($y \sim X_{\text{partisanship}}$), and compare the model fit in the two cases using the Akaike information criterion or AIC (Akaike, 1974; Stoica and Selen, 2004; Burnham and Anderson, 2004). Based on our AIC-based model selection results shown in Table D.3 and the rule-of-thumb provided in Burnham and Anderson (2004), it is clear that considering only partisanship is a significantly weaker model and considering both factors carries the 100% of explanatory or predictive power as compared to considering just the partisanship factor in our logistic regression setup. That is, considering both factors is significantly better every time for modeling co-sharing versus control than considering just partisanship as a factor.

D.3 Case Study Details: All Narrative Labels after Keyword-based Filtering

All narrative strings or labels in co-shared articles by liberal mainstream outlets that contain the terms ‘vaccin’ and ‘spread’ are included in Table D.4.

D.4 Case Study Details: Full Unique Tweet Text Set by Vaccine Misinformation Sharers

For each type 1 and type 2 mainstream article, all raw unique tweet texts that share those articles by the subset of vaccine-misinformation-sharing Twitter users (as explained in Section 5.5) are shown below:¹

‘<https://thehill.com/changing-america/well-being/longevity/561994-top-health-expert-says-vaccinated-people-are-spreading>’: [‘Top health

¹All username mentions are removed from the tweet texts and only texts with at least three space-separated tokens are retained.

Narrative Labels			
colorado pastor spread vaccine conspiracy theory	facebook announce limit reach anti vaccination content longer serve anti vaccination group page search result recommendation bar longer allow user group spread vaccine misinformation place ad run fundraiser	agency not-rule possibility fully vaccinate individual develop asymptomatic infection spread virus inadvertently urge vaccinate continue practice certain precaution	mask help prevent spread unvaccinated people school
unvaccinated trump supporter spread coronavirus conspiracy theory	group spread anti vaccine content	fully vaccinate individual spread virus	address russia struggle pandemic control kremlin spokesman dmitry peskov told rapid vaccination pace reason far unable radically contain spread disease center disease control prevention recommend american wear mask indoors high transmission area vaccinate increase spread covid delta variant
fully vaccinate people spread delta variant	recommend trump suggests use flu vaccine coronavirus caught virus recommend work told self isolate prevent far spread outbreak	crowd health worker rebut falsehood spread vaccine	initial vaccine stop spread
russia troll spread vaccination misinformation create discord	feature young pharaoh video spread conspiracy theory vaccine new world order redirects site	official believe vaccination way stop spread recognize limit effort facebook wednesday announce remove facebook instagram account largely run russia spread vaccine conspiracy theory audience primarily india latin america lesser extent united state accord report publish social medium company facebook instagram account spread vaccine conspiracy theory audience primarily india latin america lesser extent	bot russian troll spread misinformation vaccine
vaccinate people spread delta variant	vaccination aggressive test measure slow spread		bot spread malware unsolicited content disseminate anti vaccine message
unvaccinated child spread virus	state health official lead plead people vaccinate slow spread stop deadly variant develop republican control committee plan block university wisconsin institute covid test mask vaccination protocol campus state come health official sound warning rapidly spread highly contagious delta variant	vaccinate people infect delta variant vaccination spread virus	researcher conclude bot spread malware unsolicited content disseminate anti vaccine message russian troll promote discord
arrival vaccine slow spread			
advertisement hour vaccination health official release statement warn variant coronavirus continue spread la county	vaccinate spread false	previously prolific spreader vaccine misinformation told censor	vaccinate people spread disease
country figure deploy additional resource like test supply vaccine stop spread	bum explain onset case ship medical department implement health protection measure provide additional outbreak specific dose vaccine crew manage patient stop spread illness	sen amy klobuchar minn told vaccine misinformation spread online deadly consequence social medium platform action account propagate majority lie	data suggests vaccinate people spread disease
vaccine stop spread	vaccinate spread virus	previously prolific spreader vaccine misinformation tone post	advertising spread vaccine
federal official fauci public health expert argue widespread test mask wear social distance need ward far spread scientist scramble produce effective vaccine	colleague found laboratory weaken virus use oral polio vaccine rapidly regain strength start spread	effort play figure outbreak impose quarantine case try stop spread slow simultaneously work vaccine treatment	data emerge israel suggest pfizer vaccine effective infection cause delta variant explain uptick infection despite widespread vaccination
bill gyado writer caliwax madium firm active nigeria ghana daily beast recently expose russian link troll factory previously use young men woman west african nation spread disinformation democrat amplify racial issue america aim instigate social unrest write article headline africa focus russia sputnik vaccine			delta variant explain uptick infection despite widespread vaccination cause infection
vaccinate spread china	unvaccinated people spread germ	task force report caution wo late spring american vaccinate substantially low risk viral spread	irving like instagram post user claimed vaccine implant microchip plan satan baseless conspiracy theory spread nba locker room accord roll stone
workplace vaccine mandate reduce overall spread virus	social medium company spread misinformation vaccine white house spokesperson judd deere told medium democrat focus impeachment refuse seriously acknowledge virus january february president trump bold action protect american unleash power federal government curb spread virus expand test capacity expedite vaccine development true idea level transmission asymptomatic spread nbc news analysis million tweet sent identify russian troll publish data journalism website fivethirtyeight clemson university researcher found example tweet mention vaccine spread divisive misinformation discredit theory	vaccine need administer estimate percent percent population million people stop spread virus	ventilator spread vaccinate
threat argues workplace vaccine mandate reduce overall spread virus		jain taught history vaccine course stanford work publication incident add gang cage condition nt die journey sick virus spread quickly	large tech company cut spread vaccine misinformation
president spread vaccinate		biden official stress miss dos spread state remain largely responsible health provider charge vaccinate million people wait line shot	tech company crack spread misinformation surround vaccination
world health organization try set vaccine aspiration achieve widespread use	vaccine spread divisive misinformation discredit theory	merger trump cultist nationalist populism anti vax batsbittery mean develop vaccine prevent far spread disease meaningful fraction american refuse accept wo kid vaccinate	vaccine spread coronavirus
vaccinate spread medium	example tweet mention vaccine spread divisive misinformation discredit theory	vaccine prevent far spread disease	mitigation measure mask mandate help limit spread vaccinate
people understand people vaccinate vaccine wo nt help reduce spread virus	vaccine expert worry white house apply pressure fda push vaccine eua fully test pathway use approve vaccine intend widespread use	virus spread vaccinate	vaccinate people spread coronavirus
vaccine reduce spread virus	public health official administer widespread vaccination	pennsylvania medical society spread campaign debunk common anti vaccination talk point	minority community member vaccine skeptical create space highly contagious virus spread
people need understand people vaccinate vaccine wo nt help reduce spread virus	trump administration appointee argue need widespread test contain virus randomize control clinical trial determine drug vaccine work	ad spread vaccine misinformation	widespread vaccination let business usual
vaccine not-help reduce spread virus	authority allow coronavirus spread young healthy people protect elderly vulnerable approach rely arrive herd immunity infection vaccine	question shadow plan widespread vaccination	unvaccinated people help fuel rapid spread virus
president trump unleash power federal government curb spread virus expand test capacity expedite vaccine development	white house embrace declaration group scientist argue authority allow coronavirus spread young healthy people protect elderly vulnerable approach rely arrive herd immunity infection vaccine	vaccine allows disease spread faster longer normally	arrival highly effective vaccine december lift hope eventually slow stop spread disease rest population
member group spoad conspiracy outbreak preventable disease hoax perpetrate government use group mass contact parent child die suggest evidence vaccine blame	study found march early january vaccination widespread delta variant arrive proportion patient mild asymptomatic disease percent	million cite concern blood clot widespread preference pfizer vaccine	scientist not-ignore vaccinate people spread virus
user group spread vaccine misinformation	newsom supporter emphasize california state lead nation high vaccination rate low covid case rate spread delta variant decimate republican lead state florida texas south dakota	mask prevent spread unvaccinated people school	available vaccine slash odds infect people spread virus onward half likely

Table D.4: All automatically extracted narrative labels or strings in liberal co-shared articles containing the terms ‘vaccin’ as well as ‘spread’.

expert says vaccinated people are spreading delta variant — TheHill’, ‘Top health expert says vaccinated people are spreading delta variant TheHill - THAT IS LITERALLY THE HEADLINE. I wonder if President Biden will consider me a “murderer”

for sharing.', 'Top health expert says vaccinated people are spreading delta variant', 'Vaccines that target the spike protein are causing it to mutate. Virology 101. Top health expert says vaccinated people are spreading delta variant — TheHill'],

<https://www.businessinsider.com/cdc-fully-vaccinated-people-can-spread-delta-variant-2021-7>: ["BuT hoW ManY masKs sHOuld thEy WeAr?? CDC says fully vaccinated people spread the Delta variant and should wear masks: 'This new science is worrisome' via ", "CDC says fully vaccinated people spread the Delta variant and should wear masks: 'This new science is worrisome' ALL YOUR "SCIENCE" IS WORRISOME!!", 'Vaccinated and unvaccinated carry same viral loads, per the CDC.. KEEP THIS SHOW IT TO YOUR STUPID VACCINATED FRIENDS AND FAMILY, YOUR GOING TO NEED IT.', 'Who didn't see this coming. Paying attention yet?', "CDC says fully vaccinated people spread the Delta variant and should wear masks: 'This new science is worrisome' Has the CDC done anything beneficial during this entire debacle? They push masks, then the vaccine, now masks again. Fu them.", "What will tomorrow bring? A new set of opinions? CDC says fully vaccinated people spread the Delta variant and should wear masks: 'This new science is worrisome' via ", 'Science tells us vaccine resistant viruses can evolve in vaccinated groups when viruses mutate frequently, the vaccine isn't 100% effective, the virus is circulating widely, and mass vaccination is taking place. The science didn't change; the virus did.', " Well look what the CDC is saying today It's the vaccinated spreading covid."],

<https://www.npr.org/sections/goatsandsoda/2017/06/28/534403083/mutant-strains-of-polio-vaccine-now-cause-more-paralysis-than-wild-polio>: ['2 year old article is still relevant.', 'Mutant Strains Of Polio Vaccine Now Cause More Paralysis Than Wild Polio', 'WHO has reluctantly admitted the global explosion in polio is predominantly a #PolioVaccine strain. The most

frightening polio epidemics in Congo, Afghanistan, and the Philippines, are all linked to vaccines.', ' The oral polio vaccine spreads polio. Every case is caused by the vaccine. No child needs 72 doses of vaccines. Hep B vax given on day 1 of life is insane.', "None for me, thanks. I'll take my polio natural.", '#WakeUp vaccines have never been proven safe or effective.', 'Vaccine-CAUSED POLIO now more prevalent than wild caught POLIO!! When will we WAKE UP!? Vaccines cause disease!!', 'Mutant Strains Of Polio Vaccine Now Cause More Paralysis Than Wild Polio . For the first time, the number of children paralyzed by mutant strains of the polio vaccine are greater than the number of children paralyzed by polio itself.', '47. ...but with several other paralyzing viruses in existence, it is not the panacea people believe it to be. To complicate matters, the oral poliovirus vaccine itself occasionally creates paralysis.', 'NPR: Mutant strains of polio VACCINE now cause more paralysis than wild polio.', 'For the first time, the number of children paralyzed by mutant strains of the polio vaccine are greater than the number of children paralyzed by polio itself. But we're supposed to accept Gates' new #Covid19 vaccine being fast tracked right now...', ' Nope. Vaccines have been causing VARIANTS for a LONG time. All you folks that got VACCINATED are the ones who have "caused this". #TheVaccinesAreTheVariants', 'Mutant Strains of #Polio #Vaccine Now Cause More Paralysis Than Wild Polio 2017: for the first time, the # of children paralyzed by mutant strains of the polio vaccine are greater than the number of children paralyzed by polio itself. #vaccines', 'The media would never tell the public vaccines cause polio (AKA #AcuteFlaccidParalysis) Mutant Strains Of #Polio #Vaccine Cause More Paralysis Than Wild Polio', 'Too few people know that In 2017, the polio #vaccine caused more paralysis than wild #polio. #vaccines #MondayMotivation #flushot', 'Mutant Strains of #Polio #Vaccine Now Cause More Paralysis Than Wild Polio 2017: for the first time, the # of children

paralyzed by mutant strains of the polio vaccine are greater than the number of children paralyzed by polio itself. #vaccines #vexit'],

<https://www.pbs.org/newshour/science/tthis-chicken-vaccine-makes-virus-dangerous>: [" We've known since at least 2015 that an #ImperfectVaccine can make a virus worse. Repeating the same mistake is insanity.", "Just realize: 1) We've never mass vaccinated in middle of a circulating pandemic 2)We've never used such a narrow spectrum spike protein 3) We've never used such a leaky vaccine in terms of efficacy against transmission With that in mind, now read this", ' "...virus spread to sentinel birds nine days faster if it came from a vaccinated chicken versus an unvaccinated one. In addition, sentinels died faster when exposed to vaccinated chickens versus unvaccinated chickens" 'Science", 'See article for a disturbing potential parallel.', 'This chicken vaccine makes its virus more dangerous — PBS NewsHour', 'extended the lifespan of birds exposed to the hottest strains, with 80% living longer than 2 months But chickens were transmitting the virus, shedding 10,000 times more virus than an unvaccinated bird #Mareksdisease #forgottenHolo-caust #forgottenhistory', " Like Marek's disease in chickens??", 'Marek Disease!!! This chicken vaccine makes its virus more dangerous — PBS NewsHour. COVID vaccines are non-sterilizing, aka " leaky" vaccines, too. What are we doing? This cannot end well.', "Read this article on Marek's disease with a leaky chicken vaccine, and then ponder exactly what is going on here and in Israel, and ask yourself why we can rule out that this is happening to humans, now that it's confirmed pfizer is extremely leaky", 'For those interested, here's a short review.', ' Wait until he hears about (see latest tweets) and Marek's Chickens and finds out the vaxxed are the real superspreaders.'],

<https://www.washingtonpost.com/health/2021/07/30/provincetown-covid-outbreak-vaccinated>: ['Vaccinated people made up three-quarters of those

infected in a massive Massachusetts covid-19 outbreak, pivotal CDC study finds', '3/4 were vaxed in MA outbreak', 'The vaccines do not stop the spread, they barely slow it. Therefore there is zero justification for mandates.', 'This says otherwise:', 'Zero science to back up any of this. Even CDC says vaccinated still spread #COVID; mask mandated never reduced or prevented spread of #covid', 'Oh snap! Most people infected in Provincetown covid outbreak were vaccinated, CDC study shows - The Washington Post', 'Totally agree. Facts do Matter', 'Most people infected in Provincetown covid outbreak were vaccinated, CDC study shows - The Washington Post', "Wow ... looks like the CDC has got caught with its pants down! So why get the vaccine at all if it doesn't work ... and it has all those side effects??? CDC study shows three-fourths of people infected in Massachusetts covid-19 outbreak were vaccinated", 'The proper response is not to blame Blue America for outbreak but for those finger-pointing commentators to be a little more circumspect, a little less Trump-obsessed, a little more reasonable. Would be good idea. 3/3 From WP:', 'Good grief. Did not beat Covid variant. Mask up again. Booster shots? CDC study shows three-fourths of people infected in Massachusetts covid-19 outbreak were vaccinated']

D.5 Case Study Details: Measuring the Impact of Type-1 Mainstream Articles in Terms of Uptake amongst Vaccine Misinformation Sharers

By detecting if the string 'vaccin' is present in the (lowercased) headline, we find that in our final dataset of $\sim 420,000$ news articles, there are 67 vaccine-related articles published by The Hill, 47 vaccine-related articles published by Business Insider, and 444 vaccine-related articles published by The Washington Post. We

found the rank of the particular three type-1 articles (one from each of these outlets; Table 5.8) in terms of shares on Twitter amongst users who shared vaccine-related misinformation, and those who shared vaccine-related ‘real’ news (but never shared any fake news article; see the last paragraph in Section 5.5).

Using a simple normal test for proportions (proportions z test), we find that for each of the three articles, they are significantly more popular relative to all vaccine-related articles published by the corresponding outlet among the vaccine misinformation sharers ($p \lll 0.05$).

D.6 Direct Quotation Analysis: Procedure and Results

Using prior work in Poulighen et al. (2007) and some of the quotation markers present in our own dataset, we use the following markers to filter out sentences:

: , ' , " , ' , " , " , " , < < , > > , ' , '

We note that ‘:’ is included since many headlines report a direct quote as “*source name: quoted text.*”

We use the NLTK library (Bird, 2009) to split the texts into sentences, and the sentences into tokens. In order to test our function for filtering out sentences containing direct quotations, we get two different random samples of 25 news articles. We manually annotate the sentences expected to be removed via our simple marker-based filtering (ground truth). Then we run our function to automatically filter or remove the sentences, and compare them to the ground truth.

After a couple of iterations on the first random sample as a training set, we test our function on the held-out sample of 25 full article texts. Our function achieves 0.985 recall and 0.564 precision — it misses just one edge case. This is exactly what we optimized for, a very aggressive or recall-oriented function for filtering out

Type of outlets	Misinformation narrative library	Dimensionality of extracted narratives	<i>n</i>	
All reliable outlets	All fake news narratives	Low-dimensional	64318	$p < < < 0.05$
		High-dimensional	9278	$p < < < 0.05$
	Recurring fake news narratives	Low-dimensional	1836	$p < < < 0.05$
		High-dimensional	434	$p < < < 0.05$
	False-claims narratives	Low-dimensional	3474	$p > 0.05$
		High-dimensional	168	$p < < 0.05$
Only trustworthy outlets	All fake news narratives	Low-dimensional	60094	$p < < < 0.05$
		High-dimensional	15912	$p < < < 0.05$
	Recurring fake news narratives	Low-dimensional	1780	$p < < < 0.05$
		High-dimensional	393	$p < < < 0.05$
	False-claims narratives	Low-dimensional	3272	$p > 0.05$
		High-dimensional	160	$p < < 0.05$
Only liberal outlets	All fake news narratives	Low-dimensional	49160	$p < < < 0.05$
		High-dimensional	4254	$p < < < 0.05$
	Recurring fake news narratives	Low-dimensional	1546	$p < < < 0.05$
		High-dimensional	259	$p < < < 0.05$
	False-claims narratives	Low-dimensional	2911	$p < < < 0.05$
		High-dimensional	131	$p < < < 0.05$

Table D.5: Results of significance testing showing that across various categories of reliable mainstream news outlets after all sentences containing direct quotes (signified through quotation markers) were removed. Across various sets of misinformation narratives, co-shared articles have a significantly higher presence of misinformation narratives compared with the control group.

sentences containing direct quotes.

We then use this to create filtered versions of co-shared and control group of texts, and reconduct our significance tests and find that our hypothesis holds in 16 out of 18 cases (Table D.5) as opposed to 17 out of 18 cases with original, non-filtered texts (Table 5.3).

D.7 Automatically Extracted Narratives from False Fact-checked Claims

In Tables [D.6](#) to [D.26](#), we show all high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table [5.3](#)). Recall that we tested if misinformation narratives have a significantly higher presence in co-shared articles using the one-sided Wilcoxon-signed rank test (non-parametric significance test for comparing paired or dependent data) ([Wilcoxon, 1992](#); [Woolson, 2007](#)), and that before this test is performed, any occurrences (counts of narrative label in co-shared and control) of ties and zeros were removed ([Pratt, 1959](#)) (Section [5.4](#)).

Narrative label	Sentences containing narrative label
celebrity not- political	kurt russell said that celebrities shouldnt get political when it comes to donald trump gervais concluded his scathing monologue by warning the celebrities not to make any political or woke statements when accepting their awards
explosion kill miner	says obamas deadliest coverup (in an explosion that killed 29 miners) has been uncovered by internal government documents the former ceo of the coal company massey energy blankenship was convicted of conspiracy to violate minesafety laws over an explosion that killed 29 miners in 2010
company laid people	new reports show kevin nicholson made over \$1 million advising companies that shut down wisconsin plants and laid off 1900 people pepsico has kicked off a round of layoffs affecting employees in several offices two people who were laid off by the company told business insider
intel community identify alien	those are aliens who the intel community has identified are of concern in a statement last week homeland secretary kirstjen nielsen noted that us customs and border protection agents have in stopped over 3000 special interest aliens from crossing into the us those are aliens who the intel community has identified are of concern she said
american not-default tax	currently we have some 4045% of americans who are not paying any taxes also this promise falls under bidens larger proposed tax plan which stresses that americans making less than \$400000 would not pay more in taxes
jerry nadler thought concept starr report absolutely	so jerry nadler thought the concept of giving the starr report was absolutely something you could never do roy barnes is part of the team that has run up trillions of dollars in debt
american spend year	that americans will spend years paying off and theyve done it without creating one job in effect many americans spend part of the year working for nothing donating the equivalent of \$561 on average to their firms
plan parenthood sell baby	planned parenthood has now (been) found to also illegally sell baby parts kamala harris tried to put proliferers in jail who exposed planned parenthood selling baby parts it is because i dared to question planned parenthood and the abortion industry and to speak the truth: planned parenthood sells baby parts
veteran suffer ptsd	donald trump said veterans suffering from ptsd are weak unfit and cant handle military service speaking with the retired american warriors pac in virginia the republican nominee suggested veterans suffering from ptsd are not strong and cant handle the stresses of war

Table D.6: All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3).

Narrative label	Sentences containing narrative label
russian state interfere internal american affair include election process	president trump mentioned the issue of the socalled interference of russia with the american elections and i had to reiterate things i said several times including during our personal contacts that the russian state has never interfered and is not going to interfere into internal american affairs including election process it was putin not trump who pointed that out — adding: i had to reiterate things i said several times including during our personal contacts that the russian state has never interfered and is not going to interfere into internal american affairs including election process once again president trump mentioned the issue of the socalled interference of russia when the american elections and i had to reiterate things i said several times including during our personal contacts that the russian state has never interfered and is not going to interfere into internal american affairs including the election process
russian state interfere internal american affair include election process	president trump mentioned the issue of the socalled interference of russia with the american elections and i had to reiterate things i said several times including during our personal contacts that the russian state has never interfered and is not going to interfere into internal american affairs including election process it was putin not trump who pointed that out — adding: i had to reiterate things i said several times including during our personal contacts that the russian state has never interfered and is not going to interfere into internal american affairs including election process once again president trump mentioned the issue of the socalled interference of russia when the american elections and i had to reiterate things i said several times including during our personal contacts that the russian state has never interfered and is not going to interfere into internal american affairs including the election process
people use medicaid	twothirds of the people who use medicaid are poor children but twothirds of the money is for longterm care for seniors whether in a facility or at home whatever programs people attached to it in their minds ms mettlers data suggests that people who have used say medicaid unemployment insurance and veterans disability payments over the course of their lives may be more influenced by their dislike of something they havent used than by their experiences with the programs they have
terrorist attack united state	president obama said that the united states deserved being attacked by terrorists on 9/11 because we didnt respect islam but now 73% of americans believe afghanistan will once again become a base for terrorists who want to attack the united states
million people vote election	a north carolina study proves that probably over a million people voted twice in (the 2012) election when all is said and done more than 158 million people will have voted this election according to the us elections project a turnout rate of more than 66% on his sixth day in office he advanced the false claim that millions of people voted illegally in the 2016 election depriving him of a popularvote majority

Table D.7: All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); *cont.*

Narrative label	Sentences containing narrative label
country speak spanish	the chevrolet nova sold poorly in spanishspeaking countries because its name translates as doesnt go in spanish most of boyle heights residents hail from or have roots in mexico which is an important consideration when buying books because there are differences in vocabulary culture and interests among spanishspeaking countries previously rumors about the challenge spread throughout latin america and spanishspeaking countries
new coronavirus cause disease	in midmarch the nba suspended its season after a player on the utah jazz preliminarily tested positive for covid19 the disease caused by the new coronavirus there are currently no effective vaccines drugs or other immune system treatments approved to treat covid19 the disease caused by the new coronavirus in february 2020 the vatican said that pope francis and two aides had tested positive for covid19 a disease caused by a new coronavirus
golden state warrior visit white house	kellyanne conway said she would make sure that the golden state warriors were required to visit the white house after winning the nba title trump in september 2017 rescinded the golden state warriors invitation to visit the white house after they won the nba championship because stephen curry was hesitating while several players from the new england patriots who won super bowl li skipped the traditional visit
obama impose regulation	obama simply imposed regulation without proper regard to the cost side of that analysis obama did sign legislation in late 2015 ending the decadesold ban on crude oil exports but at the same time imposed regulations and pursued international policies aiming to move the world away from fossil fuels in the name of fighting global warming
taxpayer fund salary	during the 2016 campaign trump promised not to take a taxpayerfunded salary if elected president says vince polistina is collecting nearly \$70000 in taxpayerfunded salaries plus a government pension she has been paid nearly \$3 million in taxpayerfunded salary since that time the outlet reported thursday
law enforcement professional request billion	finally as part of an overall approach to border security law enforcement professionals have requested \$57 billion for a physical barrier did the democrats request steel?as part of an overall approach to border security law enforcement professionals have requested \$57 billion for a physical barrier in addition as an overall approach to border security law enforcement professionals have requested \$57 billion for a physical barrier
ammamir campa najjar infiltrate congress	ammamir campanajjar is working to infiltrate congress ammamir campanajjar is working to infiltrate congress says the narrator of mr hunters ad released on youtube in reference to the democrat running against him
people earn little	the republican authored state budget includes a provision forcing people earning as little as 12000 to buy private health insurance that could cost them as much as 4000 per year more on the dairy crisis:dairyland in distresscoronavirus and dairy:coronavirus has hit wisconsin dairy farms especially hard some farmers may even have to dump milkin the last few years thousands of wisconsin dairy farmers lost money practically every day they milked cows as an oversupplied market kept prices depressed
farmer dump milk	farmers are now dumping milk into the ditches because of a trade dispute with canada while millions of americans are worried about having enough to eat and lines at food banks grow farmers have been plowing under vegetable fields dumping milk and smashing eggs that cannot be sold because the coronavirus pandemic has shut down restaurants hotels and schools
president obama wiretapped phone	claimed new information proves he was right when he tweeted that president obama illegally wiretapped his phones during the 2016 election after all trump employed a racist conspiracy theory against barack obama which helped him gain political prominence in the republican party and later claimed that president obama had wiretapped his phones

Table D.8: All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); *cont.*

Narrative label	Sentences containing narrative label
red sox win world series	cora worked for the houston astros and cheated there to help the astro win the world series. . . then cora came to boston and cheated in boston to help the red sox win the world series i savor the shot of hedonic bliss i feel when im listening to one of my favorite bands play live hiking through a field of wildflowers in colorado or watching the red sox win the world series the red sox waited 87 long years to win the world series
court decides case	just to state this: justice kennedys son gave a billion dollar loan to trump when no one would give him a dime and justice kennedy has been ruling in favor of the trump administration position for 2 years as the court decides 54 case after 54 case the supreme courts normal process for hearing cases explainedbefore we jump into the shadow docket its helpful to understand how the court ordinarily decides cases how the court decides these cases could dramatically change the balance of power among the three branches of government thus shifting the american system of checks and balances
people commit crime	the people who committed this crime may still be at large this is to say nothing of the americans robbed or injured by people who shouldnt have been here to commit those crimes in the first place what im condemning is people who commit crime barr told the committee in our case that included hydroxychloroquine azithromycin (to prevent bacterial pneumonia from developing) zinc supplements aspirin (a regular 325 mg tablet twice a day to counter microemboli) and eventually prednisone (a steroid prescribed later in marks illness to reduce inflammation once the virus is attacking the lung)
virus attack lung	bill gates is quoted in the february 1997 issue of george magazine as saying an overpopulated planet would be choked to extinction by a lungattacking virus its quite obvious that the first chart is only counting those who actually come in with new serious cases of covid19 for example those who have trouble breathing because the virus attacked their lungs the bill eliminates reflection periods for women seeking an abortion while denying vulnerable women the full truth on what abortion means what abortion is and what abortion does in the same month otter also signed senate bill 1243 which updated the state informed consent packet to include information about the new abortion pill reversal technology for women seeking abortions meanwhile states like texas and iowa which had already put up barriers for women seeking abortions tried to use the pandemic as a cover to further erode health care access for women by claiming the procedure was nonessential a determination contradicted by leading health organizations
woman seek abortion	the flu killed 80000 people in the us last year the flu kills people about one and a half times as many children died of the 201819 flu though that flu killed 80% fewer people overall
flu kill people	says president barack obama has said that everybody should hate the police everybody hates the police
everybody hate police	sen sherrod browns vote helped pass barack obamas jobkilling health care law which slapped small businesses throughout ohio with a \$500 billion tax increase one of the dozens of attempts to repeal the affordable care act was actually called the repealing the jobkilling health care law act
health care law kill job	the senator has rejected the science showing that climate change is caused by human activity thissen asked him if he thinks climate change is caused by human activity more than 3 in 4 democratic voters 78 percent say human activity is causing climate change compared with 34 percent of republicans and 58 percent of independents
human activity cause climate change	i was the first person who ran for president without the protection of the voting rights act and i will tell you it makes a really big difference now i was the first person who ran for president in more than 50 years without the protection of the voting rights act i was the first person who ran for president without the protection of the voting rights act and i will tell you it made it makes a really big difference and it doesnt make a difference in alabama and georgia it made a difference in wisconsin where the best studies that have been done said somewhere between 40000 and 80000 people were turned away from the polls because of the color of their skin because of their age because of whatever excuse could be made up to stop a fellow american citizen from voting she added
person ran president	the white house gift shop is selling coronavirus commemorative coins the white house gift shop is selling coronavirus commemorative coins now
white house gift shop sell coronavirus commemorative coin	because the state followed president trumps cdcs guidance anyone who wants to ask why did the state do that with covid patients and nursing homes its because the state followed president trumps cdc guidance cuomo continued im not going to get into the political backandforth but anyone who wants to ask why did the state do that with covid patients in nursing homes its because the state followed president trumps cdc guidance
state follow president trump cdc guidance	

Table D.9: All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); *cont.*

Narrative label	Sentences containing narrative label
people lose job	to help people who lose their jobs amid the outbreak pelosi said the bill would strengthen unemployment insurance and boost food security initiatives like food stamps so many people have lost their jobs lowincome workers will be able to qualify for tax credits based on either their 2019 or 2020 income — a priority for democrats because people who lost jobs during the pandemic could otherwise lose out on programs like the earnedincome tax credit that require people to work to qualify for the full benefit
lettuce kill american undocumented immigrant	our rulingthe meme said lettuce killed more americans this year than undocumented immigrants lettuce killed more americans this year than undocumented immigrants last year lettuce killed more americans than undocumented immigrants so its a good thing were halting food inspections over a wall that wont work in an interview with the conservative outlet newsmx in early january for instance pompeo declared that
troop leave syria	us troops were leaving syria but that the us would keep up the fight against islamic state terrorists while also maintaining the pressure on the iranian government russia iran syria & many others are not happy about us troops leaving syria trump announced in december that us troops would be leaving syria in the coming months but has since slowed the pace of a possible withdrawal after pushback from gop lawmakers and at the same time the president took no act that in fact deprived the special counsel of the documents and witnesses necessary to complete his investigation
act deprive necessary complete investigation	advertisementbeyond this barr applauded trump for taking no act that in fact deprived the special counsel of the documents and witnesses necessary to complete his investigation under gov scott walkers publicschoo budgets a student entering kindergarten will not enjoy the same state investment in his or her education as those that came before them until they are graduating from high school people of good will hold sincere beliefs on both sides of the issue she said but maine has a vaccination optout rate that is three times higher than the national average for students entering kindergarten and the state ranks seventh in the country for the rate of nonmedical exemptions taken among schoolage children filed via the great lakes justice center for plaintiffs cheryl constantino and edward mcall the suit claims that elections officials assigned random ballots to people who had not voted yet accepted ballots after the deadline failed to verify signatures and identification of voters thousands of unsealed and unsecured ballots arrived and were all marked for biden officials instructed people to vote for democrats challenges to ballots were not recorded and poll watchers were eventually kicked out scarier still for democrats the gop is getting these votes from people who normally vote democratjohn couvillon a louisianabased pollster who surveyed the field extensively said republican gains came in districts that had long voted for democratic legislators even as they became more culturally conservative — socalled ancestrally democratic regions in rural areas north and west of lafayette we were not wasting time talking to people who werent going to vote for democrats she said
student enter kindergarten	a computer infected by malware proved a vermont power company has been targeted for disruption by russian hackers malware has also been used to quietly infect computers and then sell access to other cybercriminals who can steal banking credentials or exploit other valuable information neino said the warning says that benzene can build up inside parked automobiles and suggests that drivers roll down their windows to ventilate the cars interior prior to switching on the air conditioner the physician steps up to whatever car is idling and asks the driver to roll down the window enough for his or her arm to fit
people vote democrat	planned parenthood and the satanic temple have worked together to oppose laws restricting abortion rights in missouri the democratic attorneys general association is not dictating how democratic attorneys general deal with laws restricting abortion rights should they win office nothing in the law allows insurance companies to not only to kick you off for preexisting conditions or not accept preexisting conditions but we then put in safeguards to make sure that theres not price discrimination as a result of preexisting conditions at the time critics of the grahamcassidy bill said it would have pulled health care away from 21 million americans and allow insurance companies to deny coverage to patients with preexisting conditions maddow also played a clip of josh hawley who is running for the us senate in missouri claiming to support forcing insurance companies to cover preexisting conditions
malware infect computer	federal law requires that hospitals treat people whether or not they can pay shutting the hospitals — which treated more than 70000 people over the past 12 months — would create big holes in the states care network she said i will also renew funding — set to expire in may — for the nationwide network of hospitals that can isolate and treat people with infectious diseases and fully fund the global health security agenda so the world is ready for the next outbreak
driver roll window	
law restrict abortion right	
insurance company preexist condition	
hospital treat people	

Table D.10: All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); *cont.*

Narrative label	Sentences containing narrative label
adolf hitler never used chemical weapons (against his own people)	among those were false claims that trumps inauguration had the biggest crowds in history that the president would have won the popular vote in the 2016 if not for voter fraud that former president barack obama wiretapped trump tower and that adolf hitler had not used chemical weapons
government run health care	long waits plague patients in other countries with governmentrun health care says david jolly lobbied for governmentrun health care just last year in 2009 rock the vote launched a multimedia campaign focused on supporting the obama administrations efforts to enact governmentrun health care including a 30second television ad with celebrities encouraging young people to demand health care now
refugee enter united state	jeff foxworthy said that you might be a moron if you allow refugees to enter the united states a homeland security adviser to former vice president mike pence accused the trump administration of distorting the truth about afghan refugees writing on twitter that the former president and stephen miller his top immigration adviser sought to prevent the refugees from entering the united states washington — president biden intends to increase to 125000 the number of refugees who can enter the united states in the fiscal year beginning oct 1 the state department announced on monday making good on his campaign pledge to do so
president trump crack illegal immigration	says former democrat kelli ward attacked president trumps plans to crack down on illegal immigration saying trump was making a mistake charles mostoller for the new york timesmorristown tenn — one morning in april federal immigration agents swept into a meatpacking plant in this northeastern tennessee manufacturing town launching one of the biggest workplace raids since president trump took office with a pledge to crack down on illegal immigration
san francisco er quarterback colin	former san francisco 49ers quarterback colin kaepernick lost an endorsement deal with sports apparel manufacturer nike for going on an antiwhite rant the decision came two years after former san francisco 49ers quarterback colin kaepernick triggered a firestorm of criticism for kneeling during the anthem to protest racial injustice and police brutality
people eat bat snake dog thing like	china is to blame because the culture where people eat bats and snakes and dogs and things like that these viruses are transmitted from the animal to the people and thats why china has been the source of a lot of these viruses like sars like mers the swine flu and now the coronavirus our rulingcornyn said: china is to blame because the culture where people eat bats and snakes and dogs and things like that these viruses are transmitted from the animal to the people and thats why china has been the source of a lot of these viruses like sars like mers the swine flu and now the coronavirus sen john cornyn rtx said china is to blame because the culture where people eat bats and snakes and dogs and things like that these viruses are transmitted from the animal to the people and thats why china has been the source of a lot of these viruses like sars like mers the swine flu
supreme court upheld affordable care act	the reason that the supreme court upheld the affordable care act is because they ruled that each of these monthly payments that everyday american make is a tax and lets not forget that the reason that the supreme court upheld the affordable care act is because they ruled that each of these monthly payments that everyday american make is a tax washington (cnn) judge justin walker now nominated to a powerful federal court dubbed the second highest in the nation remains bitter that the supreme court upheld the affordable care act eight years ago
hillary clinton sell uranium	says hillary clinton presided over \$6 billion lost at the state department sold uranium to the russians through (her) faux charity illegally deleted public records and murdered an ambassador transforming politics into a posttruth contest of tribal identity is an explicit goal of modern propagandasean illingthe problem with the flood the zone strategy is that when you respond by debunking the bullshit stories — like the claim that hillary clinton sold uranium to russia — all you do is amplify it and thus help to frame the conversation on false grounds
obamacare raise premium	obamacare raised premiums by 100 percent and after that 91 percent of the people were covered but were all paying 100 percent more by raising the floor to 60 percent obamacare raised premiums by an estimated 85 percent
government run health care system	with the help of american taxpayer dollars soros bolsters a radical leftwing agenda that in the united states has included: promoting an open border with mexico and fighting immigration enforcement efforts; fomenting racial disharmony by funding anticapitalist racist organizations; financing the black lives matter movement and other organizations involved in the riots in ferguson missouri; weakening the integrity of our electoral systems; promoting taxpayer funded abortionondemand; advocating a governmentrun health care system; opposing us counterterrorism efforts; promoting dubious transnational climate change agreements that threaten american sovereignty and working to advance gun control and erode second amendment protections
kavanaugh mislead senate	says state rep patrick rose votes 90 percent with liberals backed the states largest tax increase proposed nearly \$15 billion in government spending and supported obamastyle governmentrun health care system for texas green: well to a certain extent polling backs you up on broad public support for these policies such as a governmentrun healthcare system or expanding free or lowcost public college fast forward to this week and its obvious leahy has problems with kavanaughs denials going so far as to say that kavanaugh misled the senate during his past confirmation hearings kavanaugh might have also misled the senate by denying any involvement with detainee policies or any knowledge of documents related to a warrantless surveillance program and even as evidence mounted that kavanaugh had misled the senate about his highschool behavior his drinking habits and references in his yearbook conservatives insisted democrats were raising irrelevant matters related to trivial adolescent misbehavior
democrat change law	the democrats have to change their law the democrats have to change their law thats their law the democrats have to change their law — thats their law
government sponsor enterprise	says freddie mac electric coops and credit unions are all government sponsored enterprises mainly its a sordid history of the governments seizure of mortgage giants fannie mae and freddie mac also known as the governmentsponsored enterprises or gses
police shot african american	in 2012 123 africanamericans were shot dead by police and then the following day the cover of the local newspaper would read violent africanamerican shot by police while trying to steal awardwinning pig
president obama play golf	president obama snubbed the funeral of supreme court justice antonin scalia in order to play golf can you believe that with all of the problems and difficulties facing the us president obama spent the day playing golf he notably chastised his predecessor on twitter in 2014 writing can you believe thatwith all of the problems and difficulties facing the us president obama spent the day playing golf
obama administration legalize bump stock	obama administration legalized bump stocks

Table D.11: All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); *cont.*

Narrative label	Sentences containing narrative label
fake news like talk	general flynn was given the highest security clearance by the obama administration — but the fake news seldom likes talking about that in a second tweet monday morning trump noted that general flynn was given the highest security clearance by the obama administration but the fake news seldom likes talking about that general flynn was given the highest security clearance by the obama administration but the fake news seldom likes talking about that trump tweeted monday morning
man kill wife	a man killed and ate his wife after being stuck in an elevator for 71 hours in march a man shot and killed his wife before attempting to shoot himself at the thousand oaks mall
democrat want invite caravan caravan illegal alien country	the democrats want to invite caravan after caravan of illegal aliens into our country
house impeach president	trump: says democrats want to invite caravan after caravan of illegal aliens into our country but what i would like to see is for the house to impeach the president and for the senate to convict him today its not yet clear how schiff and other potential house impeachment managers would use any new information during the trial that wasnt available when the house voted to impeach the president but house speaker nancy pelosi has touted the new development as part of her argument that the strategy of withholding the impeachment articles for nearly a month was effective the us constitution states that if a president is impeached by the house but not convicted by the senate that persons first term is nullified and they are eligible to run for office two more times
act attorney issue court order search house office	my prosecutor and the acting us attorney that issued the court orders to search my house and my office they just attended a hillary clinton fundraiser with another us attorney out of here in san diego my prosecutor and the acting us attorney that issued the court orders to search my house and my office they had just attended a hillary clinton fundraiser!
romney left massachusetts	romney left massachusetts with a \$245million debt because of the big governmentmandated health care system with ryan at his side romney who has since left massachusetts behind to move to utah and now represents utah in the us senate lost the 2012 presidential election to thenpresident barack obama
family employee pay tax legally require	says he has a fiduciary responsibility to his business his family and his employees to pay no more tax than legally required mr trump is a highlyskilled businessman who has a fiduciary responsibility to his business his family and his employees to pay no more tax than legally required the statement said
president trump propose border wall	ms waldman said president trumps proposed border wall could have helped in this case a california college student has pledged to cut off his genitals in protest of president trumps proposed border wall a central sticking point remains funding for president trumps proposed border wall and with the senate adjourned until thursday there is no apparent quick end in sight
disease threaten life	the efforts by the administration to impede this work will undermine scientific discovery and the ability to find effective treatments for serious and lifethreatening disease in poland abortion is illegal — except to save the life of the mother in cases of rape in the first trimester or if the baby has a severe and irreversible handicap or an incurable and lifethreatening disease its parents who will have to treat every sniffle or cough as a potential lifethreatening disease; its parents who will be guessing at the efficacy of expensive home rapid covid19 tests; and its parents who will be trying to juggle everything is back to normal work schedules around periodic weeklong school quarantines because some other family just had to drag their kid to church or florida or a church in florida
donald trump win presidential election	professor helmut norpoth from stony brook university told lou dobbs on august 16th 2016 that donald trump had an 87% chance of winning the 2016 presidential election wisconsin was also the state that helped donald trump win the 2016 presidential election murray said he was speaking out due to claims from intelligence officials that wikileaks was given the documents by russian hackers as part of an effort to help donald trump win the us presidential election
clapper admit spy campaign	clapper has now admitted that there was spying in my campaign on thursday trump went after clapper in a tweet again saying clapper has now admitted that there was spying in my campaign
south african government seize land	south african government is now seizing land from white farmers the south african government has begun seizing land from white farmers following an august 1 ruling which amended that nations constitution to legalize the practice says the south african government is now seizing land from white farmers and alleges a large scale killing of farmers in the country
plan increase deficit	says his plan would cut taxes without increasing the deficit republicans argued those effects were ancillary and said the plans would increase the deficit in future decades which is also a violation of budget rules
crowd chant lock	inside the rallies crowds chanted lock her up a taunt never directed at trumps male primary rivals hillary clinton refused to exit her vehicle at an event in florida after the crowd started chanting lock her up will his crowds continue to chant lock her up!
national debt hit trillion	the (national) debt really added up 75 trillion in 2010 and is on track to hit 99 trillion this year as part of a spending pattern that the cbo deemed unsustainable the national debt is expected to hit \$314 trillion by 2030 as the national debt hits \$22 trillion these are the only people trying to cut spending
people want talk mask aclu file suit	i know you people want to talk about a mask critics call it a modernday poll tax and the aclu immediately filed suit against the law after desantis signed it the aclu has filed a suit to end prayer from the military completely when the aclu filed suit to reunite them homeland security argued that there was insufficient proof that ms l was the mother and that there were fears of human trafficking
president exceed constitutional authority	the 90 decision last week was the 13th time the supreme court voted 90 that the president had exceeded his constitutional authority at the time bob goodlatte a republican representative from virginia said that the 90 decision last week was the 13th time the supreme court has voted 90 that the president has exceeded his constitutional authority

Table D.12: All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); *cont.*

Narrative label	Sentences containing narrative label
people cross border	data obtained by cnn show the exact opposite: the number of people caught illegally crossing the border has increased approximately 5% when compared to figures from april the reality is that people are being let in and crossing the border every single day media coverage of the trump administrations policy of arresting people who illegally cross the border has been almost uniformly negative
ammam campa najjar work infiltrate congress	ammam campanajjar is working to infiltrate congress ammam campanajjar is working to infiltrate congress says the narrator of mr hunters ad released on youtube in reference to the democrat running against him
democrat want caravan	they [democrats] wanted that caravan and there are those that say that caravan didnt just happen it didnt just happen the democrats want caravans they like the caravans democrats want the caravan they like the caravan he said accusing democrats of wanting to give welfare and health care to illegal immigrants
ronald reagan cut tax	after ronald reagan cut taxes in 1981 the us enjoyed exponential growth when ronald reagan wanted to increase military spending and cut taxes he didnt pay for it — he just did it
bill outlaw abortion	says mitt romney backed a bill that outlaws all abortions even in cases of rape and incest the past two weeks have been some of the worst on record for abortion rights in the us last week the governor of georgia brian kemp signed a bill that outlaws abortion after six weeks of pregnancy on friday the governor of missouri where the wellrespected school is located signed a bill outlawing abortion after eight weeks
administration accomplish administration history country	my administration has accomplished more than almost any administration in the history of our country trump told the united nations in september 2018 address in less than two years my administration has accomplished more than almost any administration in the history of our country in less than two years donald trump said at the united nations this morning my administration has accomplished more than almost any administration in the history of our country
president obama award presidential medal freedom	president obama awarded the presidential medal of freedom to weinstein weiner clinton and cosby? taylor being awarded the presidential medal of freedom by president obama in 2015 in 2011 president obama awarded bush the presidential medal of freedom
official report crime	crime in germany is up 10% plus (officials do not want to report these crimes) since migrants were accepted (on tuesday he claimed that crime was actually up despite statistics and that officials do not want to report these crimes an astonishing accusation to make without evidence for anyone other than trump who does such things regularly including making racially inflammatory fabrications)
mexico build wall	says mexico can afford to build a wall because the countrys trade deficit with america is billions of dollars back on the campaign trail trump claimed that mexico can afford to build a wall because mexicos trade surplus with america is billions of dollars trump long ago abandoned his 2016 campaign promise that mexico would pay to build a wall
obama administration grant citizenship	just out that the obama administration granted citizenship during the terrible iran deal negotiation to 2500 iranians – including to government officials just out that obama administration granted citizenship during the terrible iran deal negotiation to 2500 iranians including to government officials trump tweeted obama administration granted citizenship to 2500 iranians during nuclear deal: iran official
peer review paper	that claim is impossible to verify in the absence of a peerreviewed paper or even published data of any kind a total of 58 peerreviewed papers published in the first half of 2017 conclude that global warming is a myth breaking: early fauci emails from feb 2020 about gain of function experiments ‘abroad #faucileaks pictwittercom/mwz0dwaeof — antifabookcom (@jackposobie) june 2 2021british professor angus dalgleish best known for creating the worlds first hiv vaccine and norwegian virologist dr birger sørensen chair of pharmaceutical company immunor who has published 31 peerreviewed papers and holds several patents wrote that while analyzing virus samples last year the pair discovered unique fingerprints in the form of six inserts created through gainoffunction research at the wuhan institute of virology in china
trump build mile wall	trump is proposing to build a 2000mile wall right now supreme court allows trump to build 100 miles of wall with military funds an increasing number of proamerican groups and activists are worried that businessfirst republicans will persuade trump to abandon his wageboosting lowimmigration policies to build a few hundred miles of wall:nyt report said president donald trump quietly blocked a stealthy bridge amnesty – with advice from steve miller
richard nixon release tax return	even richard nixon released his tax returns to the public when he was running for president richard nixon released his tax returns state sen gayle goldin (d) who sponsored the bill noted that every presidential candidate since richard nixon had released their tax returns voluntarily until president trump donald trumpgop grapples with chaotic senate primary in pennsylvania trump social media startup receives commitment of billion from unidentified diverse group of investors iran thinks it has the upper hand in vienna — heres why it doesnt more ran for office
american soldier lose life	perhaps ocaiocortez didnt have any relatives to speak to who experienced these conflicts (in europe or elsewhere) but 409 399 americans soldiers lost their lives in world war ii more than 1000 american soldiers have lost their lives in afghanistan in the last 27 months hundreds of american soldiers lost their lives to the weapons that the qods force provided to its iraqi proxies
hot mic deliver box ppe	blogger says mike pence was caught on a hot mic delivering empty boxes of ppe to a nursing home and pretended they were heavy https://tco/sod5kaapk https://tco/1zodmvg2pg — david mack (@davidmackau) may 8 2020kimmels deceptively edited video went viral after it aired late thursday thanks in large part to a tweet by a pollster named matt mcdermott who wrote mike pence caught on hot mic delivering empty boxes of ppe for a pr stunt

Table D.13: All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); *cont.*

Narrative label	Sentences containing narrative label
employer cut hour	says us rep nancy pelosi said employers cutting hours is a good thing rather than lay off workers an employer cuts their hours employers agree to cut back hours anywhere from 10 to 60 percent depending on the state and workers receive a prorated share of the unemployment benefits they'd be entitled to if they were totally out of a job (so if your hours are lowered 20 percent you get 20 percent of your weekly unemployment pay)
people commit mass murder end democrat	so many of these people that commit the mass murders end up being democrats trump won there by a staggering 15 points in 2016; anthony brindisi defeated by two points a republican incumbent prone to incendiary comments like—in a radio interview shortly after the parkland school shooting—its interesting that so many of [the] people that commit the mass murders end up being democrats
senator ditch cocaine mitch	one of my goals as us senator will be to ditch 'cocaine mitch one of my goals as us senator will be to ditch cocaine mitch
mr assange told couple day massive dump email damage hillary clinton campaign	mr stone told mr trump that he had just gotten off the phone with [wikileaks founder] julian assange and that mr assange told mr stone that within a couple of days there would be a massive dump of emails that would damage hillary clintons campaign mr stone told mr trump that he had just gotten off the phone with julian assange and that mr assange told mr stone that within a couple of days there would be a massive dump of emails that would damage hillary clintons campaign mr stone told mr trump that he had just gotten off the phone with julian assange and that mr assange told mr stone that within a couple of days there would be a massive dump of emails that would damage hillary clintons campaign
insurance company deny coverage	before the aca insurance companies routinely denied coverage to people who became sick while they were not covered or people with a medical condition who lost coverage introduced last year the protect act would prohibit insurance companies from denying coverage to people on the basis of preexisting conditions and would also prohibit them from charging certain populations higher premiums the medicare proposal by rep paul ryan rws would allow insurance companies to deny you coverage and drop you for preexisting conditions
restriction distance social	read: georgias experiment in human sacrificemeanwhile most states have begun lifting the socialdistancing restrictions that had temporarily slowed the pace of the pandemic creating more opportunities for the virus to spread a photograph shows protest signs carried by demonstrators calling for the loosening of covid19 socialdistancing restrictions cops on the scene capriciously decided that this transgressed the states socialdistancing restrictions
bernie sander won democratic nomination	an analysis showed that bernie sanders would have won the democratic nomination if it were not for the super delegates an analysis showed that bernie sanders would have won the democratic nomination if it were not for the super delegates trump tweeted on the eve of the convention
libya reduce payment owe family american kill libyan terrorism	purevals lobbying firm made millions helping libya reduce payments owed to families of americans killed by libyan terrorism the ad accuses the candidate of selling out americans explaining that purevals lobbying firm made millions helping libya reduce payments owed to families of americans killed by libyan terrorism
republican try repeal affordable care act	as republicans try to repeal the affordable care act they should be reminded every day that 36000 people will die yearly as a result andy slavittopinion columnistearlier this month over a year after republicans tried multiple times to repeal the affordable care act i asked people in the twittersphere if their representatives in congress had voted for repeal and if so if they held a town hall to explain their vote and put forward a better vision for health care
muslim migrant beat dutch boy	muslim migrant beats up dutch boy on crutches! jayda fransen @jaydabf video: muslim migrant beats up dutch boy on crutches!
new health care insurance option low premium	as a candidate i promised that we would protect coverage for patients with preexisting conditions and create new health care insurance options that would lower premiums in fact its so dishonest that some clever editor appears to have subversively snuck links into the text that debunk some of its key claims — its hard to believe that trump or his communications staff would have done soas a candidate i promised that we would protect coverage for patients with preexisting conditions and create new health care insurance options that would lower premiums
nordstrom treat daughter ivanka steel announce building new steel mill	are democrats uniting around medicare for all?in his editorial trump also claimed: as a candidate i promised that we would protect coverage for patients with preexisting conditions and create new health care insurance options that would lower premiums my daughter ivanka has been treated so unfairly by @nordstrom us steel just announced that they are building six new steel mills us steel just announced that they are building six new steel mills trump said july 31 in tampa our rulingtrump said us steel just announced that they are building six new steel mills
democrat misstate medicaid white house cooperate special counsel investigation	democrats purposely misstated medicaid under new senate bill — actually goes up the white house fully cooperated with the special counsels investigation providing unfettered access to campaign and white house documents directing senior aides to testify freely and asserting no privilege claims nonetheless the white house fully cooperated with the special counsels investigation
trump won popular vote	final election 2016 numbers: trump won the popular vote more than a dozen states where trump won the popular vote filed briefs in support of texas action as did over 120 republican members of congress including house minority leader kevin mccarthy rcalif in friend of the court briefs he believes this is necessary because of a long long laundry list of conspiracy theories that he believes because see above and also because if you take out those two states of california and new york trump would have won the popular vote
state expand medicaid	states that expanded medicaid tend to see their populations having greater access to a wide array of health care services from primary care to substance abuse treatment to surgery governing magazine has estimated that in the states that dont expand medicaid a minimum they said of 40 percent of rural hospitals will close in those states the strange glitch exists because the law originally required all states to expand medicaid and thus did not set up a system of subsidies for the poorest americans to buy private coverage
president trump mention issue interference russia american election	president trump mentioned the issue of the socalled interference of russia with the american elections and i had to reiterate things i said several times including during our personal contacts that the russian state has never interfered and is not going to interfere into internal american affairs including election process once again president trump mentioned the issue of the socalled interference of russia when the american elections and i had to reiterate things i said several times including during our personal contacts that the russian state has never interfered and is not going to interfere into internal american affairs including the election process
parent change diaper	sexuality educator deanne carson said parents should ask a babys permission before changing their diaper parents should ask their babies permission before changing their diapers — in order to advance a culture of consent in the home a sexuality and relationships expert said
woman vote donald trump	hillary clinton said women voted for donald trump because they caved to pressure from the men around them women are coming out to vote against donald trump because they see him as a culturalformer gop governor christine todd whitmann has echoed these sentiments on msnbc these women voted for donald trump in 2016
company laid worker	says opponent bill flores helped run a company that laid off over 3000 workers and paid off executives with millions many companies have said their ceos and other top executives would take pay cuts in 2020 as the coronavirus pandemic slashed financial forecasts and companies furloughed or laid off workers
allergy sufferer find relief	the virtual shutdown of the economy led to a staggering surge in unemployment claims as many companies have laid off workers because of a halt to business allergy sufferers find relief in immunotherapy toothpaste many allergy sufferers also find relief in overthecounter products such as the neti pot and other nasal rinses for sinus problems
democratic control congress	democrats hold 8point advantage in congressional preferencethe earlier nbc/wsj poll — conducted aug 18 through aug 22 — showed democrats with an 8point lead in congressional preference with 50 percent of voters preferring a democraticcontrolled congress and with 42 percent wanting republicans in charge despite that economic optimism however the poll shows that democrats enjoy a 10point advantage on congressional preference with 50 percent of registered voters wanting a democraticcontrolled congress versus 40 percent who want a gopcontrolled one the democraticcontrolled congress however with passage of the equality act would gut the protection of religious liberty

Table D.14: All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); *cont.*

Narrative label	Sentences containing narrative label
teenager buy glock	obama told the audience it is easier for a teenager to buy a glock than to get his hands on a computer or even a book we flood communities with so many guns that it is easier for a teenager to buy a glock than get his hands on a computer or even a book in july 2016 obama asserted: we flood communities with so many guns that it is easier for a teenager to buy a glock than get his hands on a computer or even a book
plan parenthood oppose pro choice justice	i would note that planned parenthood opposed three prochoice justices just because they were nominated by republican presidents david souter sandra day oconnor and justice kennedy i would note that planned parenthood opposed three prochoice justices just because they were nominated by republican presidents: david souter sandra day oconnor and justice kennedy collins said recalling the organizations opposition to justice anthony kennedy
grandparent immigrate america	i have grandparents who immigrated to america legally they did it the correct way and its time we uphold our laws and get this wall built! says all my grandparents immigrated to america i have grandparents who immigrated to america legally they did it the correct way and its time we uphold our laws and get this wall builtits up to americans to help out and pitch in to get this project rolling
million american lose job	speaking to times senior white house correspondent brian bennett as part of the time100 talks series on tuesday kushner said theres risk in anything but the president carries the burden of the 30 million americans who have lost their jobs due to this historic effort to save lives millions of americans have lost their jobs after dr birx and dr fauci sat with the president the us went into lockdown and since that time over 30 million americans have lost their jobs and thousands of companies have gone bankrupt
officer kill alton sterling	the officer who shot and killed alton sterling was photographed at a kkk rally in 2018 it wasnt enough to charge the officers who killed alton sterling
illegal alien murder american	in 2017 illegal aliens murdered 1800 americans too many americans have been murdered by illegal aliens and too many illegals are taking advantage of the united states taxpayers with no means of ever contributing to our society kolage wrote on the gofundme page illegal aliens murdered 1800 americans mcclintock said
mmr vaccine cause autism	courts have quietly confirmed that the mmr vaccine causes autism andrew wakefield the british doctor disciplined for fraud over his claim that the mmr vaccine caused autism attended trumps inauguration
facebook user share post	croger is giving away free grocery coupons or gift cards to facebook users who like and share a post the facebook user who shared the post did not return a request for comment facebook users can receive a \$150 coupon for dollar general by liking and sharing a post
democrat link firm fusion gps	the democratlinked firm fusion gps actually took money from the russian government while it created the phony dossier thats been the basis for all of the russia scandal fake news (trump tweet july 29 2017) the democratlinked firm fusion gps actually took money from the russian government while it created the phony dossier thats been the basis for all of the russia scandal fake news
obama sent million	says obama sent \$450 million to china to build a wind farm in texas report: obama quietly sent \$221 million to the palestinian authority hours before leaving office
plan parenthood provide service	as lifenews reported at the time chelsea revealed that her mchadimred maternal grandmother was the child of unwed teenage parents who did not have access to services that are so crucial that planned parenthood helps provide in addition the other services that planned parenthood claims to provide such as prenatal services (9055) and adoption referrals (2831) account for a measly 01% of all services the report also shows that over time despite the fact that planned parenthood is providing fewer services for fewer clients at fewer facilities it has received ever increasing amounts of taxpayer funding and its political arms are spending record amounts of money on elections while fridays order is a significant victory for data recipients and prospective petitioners republicanled states are currently asking us district judge andrew hanen to declare the program unlawful and terminate it
republican lead state	idaho was the only republicanled state in the nation that chose to partner with president obama by creating a statebased marketplace under the affordable care act the michigan secretary of state has sent ballot applications — not the ballots themselves — to registered voters a growing practice among election officials including in states led by republicans
horrible law require child separate parent cross border	a horrible law requires that children be separated from their parents once they cross the border into the us our rulingtrump said a horrible law requires that children be separated from their parents once they cross the border into the usthere is no such law
people talk mask trump campaign interfere election	i know you people want to talk about a mask #fisamemo shows real collusion between dem operatives & key officials at the fbi & doj to spy on the #trump campaign & interfere in the 2016 election the dossier has been a source of intense controversy largely due to it being one of the primary sources for the nowdebunked allegations that the trump campaign conspired with russia to interfere in the 2016 election these discrepancies — and the meeting itself — have since become a key part of special counsel robert muellers investigation into whether the trump campaign colluded with russia on its attempts to interfere in the 2016 us election
government fund care	says mitch mcconnell was cured of polio when he was a child by governmentfunded care and now hes trying to take that same care away from millions of americans but linda j blumberg a health policy fellow at the urban institute argued that a single payer system such as the one proposed by senator bernie sanders of vermont could indeed create strong incentives for people with serious health problems to enter the country or remain longer than their visas allow in order to get governmentfunded care
parent drive child	half of children struck by cars near schools are hit by parents driving children to school the prosecution disagrees arguing genital mutilation is an illegal secretive and dangerous health care service that involves interstate commerce on a number of fronts: text messages are used to arrange the procedure; parents drive their children across state lines to get the procedure; and the doctor uses medical tools in statelicensed clinics to perform the surgeries
rule prevent child	this rule could prevent children under 18 from using such tools as a power screwdriver a milking machine or something as simple as a wheelbarrow on the family farm this rule would take food away from families prevent children from getting school meals and make it harder for states to administer food assistance the michigan senator added
russia not-meddle election	i never said russia did not meddle in the election while in japan for the g20 summit a reporter asked trump on friday if he would tell russia not to meddle in us elections i strongly disagree w/ statement that russia did not meddle in 2016 election
senate not-acquit president	the us constitution states that if a president is impeached by the house but not convicted by the senate that persons first term is nullified and they are eligible to run for office two more times i believe that a president cant be removed from office if there is no reasonable possibility that the senate wont convict and remove the president then the house should not be impeaching the president in the first place
mexico pay wall	mexico is paying for the wall through the many billions of dollars a year that the usa is saving through the new trade deal the usmca that will replace the horrendous nafta trade deal a man runs for president on a platform of building a wall that he says mexicos going to pay for
bunch computer model attempt tell happen year employee speak english	climate change is nothing but a bunch of computer models that attempt to tell us whats going to happen in 50 years or 30 the obama administration is suing a wisconsin company for forcing employees to speak english it would be a violation of federal antidiscrimination law for an employee to be required to only speak english except for in a situation where it was necessary for someone to do their job according to the equal employment opportunity commission
summary trigger lot discussion analysis await come	i think — i think — i suspect that they probably wanted more put out but in my view i was not interested in putting out summaries or trying to summarize because i think any summary regardless of who prepares it not only runs the risk of you know being underinclusive or overinclusive but also you know would trigger a lot of discussion and analysis that really should await everything coming out at once i suspect that they probably wanted more put out but in my view i was not interested in putting out summaries or trying to summarize because i think any summary regardless of who prepares it not only runs the risk of who prepares it not only runs the risk of you know being underinclusive or overinclusive but also you know would trigger a lot of discussion and analysis that really should await everything coming out at once he added but in my view i was not interested in putting out summaries or trying to summarize because i think any summary regardless of who prepares it not only runs the risk of you know being underinclusive or overinclusive but also you know would trigger a lot of discussion and analysis that really should await everything coming out all at once i do not say that climate change is a hoax perpetrated by the chinese trump has long doubted climate change and once called it a hoax perpetrated by the chinese
chinese perpetrate hoax	

Table D.15: All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); *cont.*

Narrative label	Sentences containing narrative label
delivery boy cause tremendous loss	unlike others they pay little or no taxes to state & local governments use our postal system as their delivery boy (causing tremendous loss to the us) unlike others they pay little or no taxes to state & local governments use our postal system as their delivery boy (causing tremendous loss to the us) and are putting many thousands of retailers out of business!
trump administration sent letter	the trump administration sent a letter to recipients of the presidents education award misspelling the word success mitchell: the ambassador also said the trump administration sent a letter to iran last night and iran has responded but neither side will provide details about those letters meanwhile sen rick scott rfla urged the trump administration to declare a public health emergency and sent a letter to the cdc requesting information about the agencys plan to combat the virus
epa regulate greenhouse gas	a bill that would halt the epa from regulating greenhouse gases would help stop rising gas prices the obama epa rule was crafted after the landmark 2007 supreme court decision that said epa has authority under the clean air act to regulate greenhouse gases though it stopped short of laying out just how far epa could go toward curbing the pollution that scientists warn will need to drop to near zero within a few decades to stave off the worst effects of climate change
house republican repeal obamacare	the epa is still required to regulate greenhouse gases but the cpps new replacement the affordable clean energy (ace) rule is drastically weaker says house republicans spent \$50 million on 33 votes to repeal obamacare health insurer aetna accidentally disclosed giving aan \$3 million in 2011 when house republicans pushed to repeal obamacare during their first year back in the majority
legislature expand medicaid	if the legislature refuses to expand medicaid 27 states are going to get our money – virginia taxpayer money – into their states to provide health care for their citizens paid for by us after highlighting her legislative successes alvarado talked about the bills that would have been vital for her district if her republican colleagues had not blocked them: one in five residents of harris county are uninsured but the legislature refused to expand medicaid; texas has some of the lowest voter turnout rates in the country but it wont adopt online voter registration; despite the recent mass shootings in el paso and santa fe alvarado couldnt pass a single one of her gun control bills
wind stop blowing	and when the wind stops blowing it doesnt make any difference does it? when the wind stops blowing thats the end of your electric
program walk gun	did the wind stop blowing id like to watch television today guys? the justice departments fast and furious gunwalking program started in the bush administration the most recent example was in the fast and furious scandal named for an obamaera gunwalking program overseen by the bureau of alcohol tobacco firearms and explosives that left a border patrol agent named brian terry dead
drug fight virus	it is an antiviral drug a drug that fights viruses cocaine kills corona virus scientists is shocked to discover that this drug can fight the virus if all goes well molecules like these could join the growing arsenal of public health measures and drugs already in place to fight the virus she said: this is another tool you could have
insurer raise premium	the republicans alternative health care plan will still let insurers raise premiums four times faster than wages in states with a community rating waiver insurers would have been able to raise premiums for one year for people with preexisting health conditions who did not have coverage for at least 63 straight days the previous year the reasoning is that any expensive treatments related to coronavirus will be offset by the delay or cancellations of elective procedures—and if they arent insurers will just raise premiums next year and americans will continue to have to pay them no matter what because they dont want to die
federal government provide relief	mitch mcconnell said he wanted states to go bankrupt instead of the federal government providing relief in a may letter to local congressional members members of las black community urged the federal government to provide relief to renters homeowners and good momandpop landlords while taking steps to stop home flippers and large investment companies from scooping up distressed properties as they did in the wake of the great recession
donald trump won presidential election	he predicts a wave of restaurant bankruptcies unless the federal government provides more relief the day donald trump won the us presidential election i found myself gripped by a profound sense of dread the southern district has been targeting conservatives and trump supporters since donald trump won the 2016 presidential election donald trump won the 2016 presidential election because he won swing states ohio pennsylvania michigan wisconsin and florida all of which have large populations of gun owners
trump play golf	trump who was playing golf at the time fox news and other networks called the race for biden issued a statement saying biden should not be declaring victory every time trump comes here to play golf florida taxpayers are paying for it read more: trumps golf outings have reportedly cost taxpayers more than \$100 millionthe white house hasnt explicitly acknowledged that trump was playing golf or who his partners were during his visit to the club
federal government spent privilege rent donald trump golf cart protect donald trump round	the federal government has spent over \$150000 just the secret service alone for the privilege of renting donald trump golf carts to protect donald trump on his rounds
republican support amendment	says a photograph of harriet tubman shows she was a guntotting democratsupporting 2nd amendmentsupporting republican in this case the motion would have removed an amendment—supported by some republicans—to exclude cases of rape and incest from an abortion ban that had already passed the house
illegal immigrant rap child	some 108000 us children were raped by illegal immigrants in 2018 previously deported illegal immigrant who raped child after philadelphia release pleads guilty to reentrytelling added while we may live in a time of heated political rhetoric each of us are responsible for setting limits on political speech to not put peoples lives at risk previously deported illegal immigrant who raped child after philadelphia release pleads guilty to reentry
president pass legislation	rare is the president whos passed more legislation whos done more things than what weve done between the executive orders and the jobkilling regulations that have been terminated theres never been a president whos passed more legislation trump said five months after taking office exempting only fdr
people count vote	this is the 21st century version of stalins ruthless observation its not the people who vote that count its the people who count the votes the people who count the votes decide everything however pick said four people stay behind and continue counting and tabulating well into the night including the woman with blonde braids continuing to count votes based on the surveillance footage
individual business pay tax	it could double the amount of taxes every individual and business pays ending medicare as we know it mnuchin said the delay would free up \$300 billion of liquidity in the economy as individuals and businesses had more time to pay their taxes
president barack obama sign executive order	when president barack obama signed the executive order (on daca he) actually said he doesnt have the right to do this the cias torture program was shut down by an executive order signed by president barack obama in january 2009 but the physical legacy of torture is on display in guantánamos courtroom intelligence is allocated to a specialaccess program when the vulnerability of or threat to specific information is exceptional according to executive order 13526 which was signed by us president barack obama in late 2009 and details how to properly handle and protect classified nationalsecurity information says charlie bass supports paul ryan plan that forces seniors to pay \$6400 a year more for health care so millionaires can pay less in taxes
millionaire pay tax	during the 2020 primary the vermont senator penned a fiery letter to the cap board accusing tanden of maligning my staff and supporters and belittling progressive ideas after its nowdefunct website thinkprogress produced a video suggesting he had softened his demand for millionaires to pay more in taxes after becoming one himself says the state auditor found that 37 percent of the 428000 students receiving free and reducedprice lunches are ineligible
student receive free reduce price lunch	eight in 10 students receive free or reducedprice lunches
large audience witness inauguration period	that was the largest audience to witness an inauguration period now federal investigators and congress are taking a look at a controversy surrounding the presidents very first day in office this was the largest audience to ever witness an inauguration period at about 540pm that day spicer began a now notorious press briefing at the white house in which he falsely stated: this was the largest audience to ever witness an inauguration – period

Table D.16: All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); *cont.*

Narrative label	Sentences containing narrative label
drug steal life	mothers and children trapped in poverty in our inner cities . . . and the crime and the gangs and the drugs that have stolen too many lives and robbed our country of so much unrealized potential
people lose family member	his american carnage inauguration speech (which in this company looks pretty tame) january 21 2017:the crime and the gangs and the drugs that have stolen too many lives and robbed our country of so much unrealized potential there were stories from people who have lost family members to covid19 when us president donald trump was asked what hed say to people who lost family members to covid19 he replied i would say this didnt hurt anyone as much as it hurt me people have lost family members are sick themselves or lost their jobs
men face charge	three other men are currently facing charges chris jacobs has been financed by the same men who are now facing charges in one of western new yorks largest public corruption schemes
company sell gun	episode 6: legaciesavailable now: three men faced charges in the murder of arizona journalist don bolles you may wonder why a company that doesnt manufacture or sell guns is wading into this issue but for us its simple bergh wrote if the company sold guns to employees or required them to buy their own weapons for work g4s managers would lose their ability to control what workers do with the weapons he said suntrust bank cancels account because company sells guns
kavanaugh chooses word	kavanaugh chooses his words very carefully and this is a dog whistle for going after birth control harris tweeted kavanaugh chooses his words very carefully and this is a dog whistle for going after birth control kavanaugh chooses his words very carefully and this is a dog whistle for going after birth control the senator tweeted
citizen carry gun	everywhere that we have more citizens carrying guns crime is less
immigrant cross mexican border	november 18 2014: russia changes its laws to allow citizens to carry guns in public for selfdefense thanks to the immigrants who illegally cross the us mexican border and the democrats who refuse to stop them the measles virus has been declared a public health emergency in 2019 one of his most pointed messages concerned president donald trumps zerotolerance immigration policy in which us authorities plan to criminally prosecute all immigrants caught crossing the mexican border illegally holding adults in jail while their children are sent to government shelters
congress cut funding	congress has cut funding has slashed funding for veterans benefits over these last years the national association of secretaries of state (nass) beginning in 2005 passed resolution after resolution requesting that congress cut off its funding
doctor mother determine execute baby	and then the doctor and the mother determine whether or not they will execute the baby they wrap the baby beautifully and then the doctor and the mother determine whether or not they will execute the baby then the doctor and mother determine whether or not they will execute the baby
iran obtain nuclear weapon	secret documents reveal that iran could obtain a nuclear weapon far sooner than we were told trump said any country that helps iran obtain nuclear weapons would also be strongly sanctioned trump is convinced that the government of iran is secretly violating the terms of the deal and that despite the rigid inspection regime the deal is actually helping iran obtain a nuclear weapon
legislature refuse expand medicaid	if the legislature refuses to expand medicaid 27 states are going to get our money – virginia taxpayer money – into their states to provide health care for their citizens paid for by us after highlighting her legislative successes alvarado talked about the bills that would have been vital for her district if her republican colleagues had not blocked them: one in five residents of harris county are uninsured but the legislature refused to expand medicaid; texas has some of the lowest voter turnout rates in the country but it wont adopt online voter registration; despite the recent mass shootings in el paso and santa fe alvarado couldnt pass a single one of her gun control bills
doctor ask gun	medicare regulations require that doctors ask patients whether they own guns dr przebinda argues that doctors should almost never ask their patients about guns except in very rare very exceptional circumstances — for example if a patient is despondent or homicidal
border city el paso texas consider nation dangerous city	the border city of el paso texas used to have extremely high rates of violent crime — one of the highest in the country and considered one of our nations most dangerous cities the border city of el paso texas used to have extremely high rates of violent crime — one of the highest in the country and considered one of our nations most dangerous cities trump said in his state of the union address tuesday the border city of el paso texas used to have extremely high rates of violent crime— one of the highest in the country and considered one of our nations most dangerous cities
bernie sander drop presidential race	wikileaks founder julian assange said bernie sanders was threatened to drop out of the presidential race when bernie sanders dropped out of the presidential race reade took to twitter to lament the news:many of you that supported different dem candidates are struggling right now a woman who has accused joe biden of sexual assault said that there is no democracy for her after bernie sanders dropped out of the presidential race on wednesday
president cut billion	says barack obama is the only president to ever cut \$500 billion from medicare the president who unsuccessfully advocated cutting \$45 billion from the nih budget last year tweeted that us cancer death rates are the lowest in recorded history adding there is a lot of good news coming out of this administration
american lose life	every day 100 americans lose their lives to gun violence eventually the last monday in may was declared a federal holiday to pay respects to all americans who lost their lives during any of the countrys wars youve got 200000 to 2 million americans are losing their lives before the fall
senator bipartisan let know russian record	two senators — bipartisan — reached out to the election apparatus in florida to let them know that the russians are in the records and all they have to do if those election records are not protected is to go in and start eliminating registered voters this is no fooling time and thats why two senators bipartisan reached out to the election apparatus of florida to let them know the russians are in your records and all they have to do if those election records are not protected is to go in and start eliminating registered voters nelson said wednesday
straight year major hurricane strike land united state bore hole	says 11 straight years of no major hurricanes striking land in the united states bores a hole right through the whole climate change argument 11 straight years of no major hurricanes striking land in the united states bores a hole right through the whole climate change argument
president trump order fbi	on friday president trump ordered the fbi to conduct a limited supplemental background investigation into the allegations as the senate judiciary committee voted to endorse kavanaugh in a 1110 partyline vote president trump ordered the fbi to raid cdc headquarters in atlanta in hopes of uncovering data that links childhood vaccinations to autism the disclosure by montgomery county police chief j thomas manger and montgomery county states attorney john mccarthy came friday the same day that president trump ordered the fbi to conduct a limited supplemental investigation

Table D.17: All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); *cont.*

Narrative label	Sentences containing narrative label
vaccine cause autism	renée directa: the conspiracies are coming from inside the houseby 2010 the majority of broadcast and print media had stopped covering discredited claims that vaccines caused autism as the video shows below healthy triplets all became autistic within hours of vaccination once again demonstrating that vaccines cause autism some of those include opposing vaccinations saying vaccines cause autism and that childrens programming normalizes autism
million million illegal vote cause lose popular vote	between 3 million and 5 million illegal votes caused him to lose the popular vote
teacher lose job	between 3 million and 5 million illegal votes caused me to lose the popular vote 3000 teachers lost their jobs when (cris) was governor he heard from other teachers who had gone through similar leaves of absence and teachers who lost their jobs after becoming sick and spouses of teachers who died after getting sick
bernie sander won nevada	say bernie sanders won nevada after all when bernie sanders won nevada the most diverse state to have voted so far at that point he got chris matthews breaking down about how theres going to be a communist revolution in the street and people are going to cut off his head in central park
people vote donald trump	archived recording 1 so people are coming here to vote for donald trump the people who voted against donald trump may have lost at the ballot box but they can win at the cash register coulter said thirty percent of the people who voted for donald trump had voted for president obama
obama administration gitmo release vicious prisoner	122 vicious prisoners released by the obama administration from gitmo have returned to the battlefield
beach friend encourage ford	march 7 122 vicious prisoners released by the obama administration from gitmo have returned to the battlefield we learned last week that a woman named monica mclean was ms fords roommate and she was one of the so-called 'beach friends' that encouraged ms ford to go to [sen] we learned last week that a woman named monica mclean was ms fords roommate and she was one of the so-called beach friends who encouraged ms ford to go to dianne feinstein and the partisan democrats on the judiciary committee
obama administration leak information	the obama administration leaked information deliberately or otherwise that led to the identification of the pakistani doctor that helped us in achieving our goals and killing bin laden the obama administration secretly recorded the phone calls illegally unmasked flynn's name and then illegally leaked the information to the media
president obama separate child	president obama separated children president obama separated the children ... just so you understand — president obama separated the children
people not- fishing	people cant go fishing anymore because of obama
illegal immigrant commit crime	he added that ice could prevent numerous crimes if local law enforcement complied with detainees and asked what those jurisdictions to say to the family of a victim of a crime committed by an illegal immigrant ralph northam cast the deciding vote in favor of sanctuary cities that let illegal immigrants who commit crimes back on the street no family should have to go bury their 20-year-old son because an illegal immigrant was allowed to sneak in this country commit crimes in this country and stay in this country
state department allow personal email account	when i got to work as secretary of state i opted for convenience to use my personal email account which was allowed by the state department march 10 2015 – clinton holds a 20-minute press encounter at the united nations saying: at the end i chose not to keep my private personal emails – emails about planning chelseas wedding or my mothers funeral arrangements condolence notes to friends as well as yoga routines family vacations the other things you typically find in inboxes ... i opted for convenience to use my personal email account which was allowed by the state department because i thought it would be easier to carry just one device for my work and for my personal emails instead of two
police kill white	same year 326 whites were killed by police according to an aclu analysis between 2013 and 2020 a black person was 16 times more likely than a white person to be killed by police in riverside county an alarming finding that confirms the lived experience of police brutality reported by our black community members their letter stated
government run health care program	the new health care law will force seniors into barack obamas governmentrun health care program michael barbarobut what about the people who are being covered by governmentrun health care programs like medicare? and weirdly the white house report takes a moment to acknowledge this noting that medicare (a governmentrun health care program that sets health care prices) has lower administrative costs than private insurers
coronavirus cause disease	new york citys schools are among thousands of schools and childcare facilities across the united states that will be closed this week: as of last night more than half of the states had announced statewide school closures due to concerns about the spread of covid19 the disease caused by the coronavirus its a daunting operation but one that public health officials say is needed alongside testing in order to know where covid19 the disease caused by the coronavirus has been spread in a community and if the cases are few enough that stay-at-home orders can be lifted and businesses reopened this week she met with drs fauci stephen m hahn the commissioner of the food and drug administration and robert r redfield the director of the centers for disease control and prevention where the group talked through concerns about the trump administrations embrace of malaria drugs as possible treatments for covid19 the disease caused by the coronavirus according to a senior administration official because of that reporting he stopped using that phone and the country lost valuable intelligence
country lose valuable intelligence	during the democratic debates in june harris also raised her hand when the moderator asked candidates about whether theyd be willing to abandon private insurance in favor of a governmentrun plan on the issue of health insurance do you believe like a number of democratic party hopefuls that private insurance should be banned in the united states and that americans should be forced into a governmentrun plan? on health insurance do you believe like a number of democratic party hopefuls that private insurance should be banned in the united states and americans should be forced into a governmentrun plan?
government run plan	
black kill percent white homicide victim	says crime statistics show blacks kill 81 percent of white homicide victims when trump began running for president in 2015 he also tweeted a meme spreading the lie that 81 percent of white homicide victims are killed by blacks
business hit record high	did small business optimism just hit a record high? research from the progressive policy center center for american progress shows businesses used cash saved from the tjia to buy back stocks hitting record highs
company donate money	at least 18 companies have donated or pledged to donate money to the blm global network foundation according to a list compiled by the washington-based capital research center which monitors nonprofits and charities boycott these 60 companies that donate money to fund the systematic murder of unborn human babies not only did he give the public supermarket chain permission to be the sole retail distributor for the vaccine in florida (after the company donated money to him) but he also came under fire for setting up a popup vaccination clinic in manatee county which restricted access to people from only the two most affluent neighborhoods in the county
people enforce law	in other words dont blame the criminals blame the laws theyre breaking and the people enforcing those laws tell them to change the laws but dont get mad at the people who enforce the laws even the employees union for the irs the very people in charge of enforcing this law are begging to be let out from under this law

Table D.18: All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); *cont.*

Narrative label	Sentences containing narrative label
fbi spy trump campaign	specifically he has said he wants to review whether the fbi had appropriate predicate for spying on the trump campaign for example when it sought a surveillance warrant against former campaign adviser carter page the fbi always intended to spy on the trump campaignwhen news first broke that the obama administration had obtained a fisa order to surveil page democrats and the leftleaning press argued the fbis surveillance of the former trump foreign policy advisor didnt constitute spying on the trump campaign because the courtordered surveillance didnt begin until after page had left the campaign perhaps this explains why the media is so desperately invested in obscuring the fact that the fbi spied on the trump campaign using russian disinformation
genius found country	a list of a country founded by geniuses but run by idiots entries originated with comedian jeff foxworthy sen kennedy: 'this country was founded by geniuses but is being run by a bunch of idiots
people seek asylum	but the problems on the ground in places like tijuana have nothing to do with physical security and everything to do with a trump administration policy aimed at deterring people from seeking asylum but its also not clear that the trump administration is interested in any solution that will make it easier for people to seek asylum legally in the us during the first democratic debate—in june 2019 when the last border crisis was in full swing—biden in a moment of honesty said that if he becomes president there should be a surge to the border that all those people seeking asylum they deserve to be heard and that if you want to flee and you are fleeing oppression you should come
people find offensive	while zuckerberg was at times conciliatory he was defensive of facebooks stance not to make what he views as kneejerk decisions against content that people could find personally offensive the claim: attorneygeneral george brandis says people have a right to be bigots and to say things that other people find offensive what we will do he said is well say okay you have your page and if youre not trying to organize harm against someone or attacking someone then you can put up that content on your page even if people might disagree with it or find it offensive
rosie o'donnell commit offense	rosie odonnell apparently committed the same offense five times five times when she broke the identical law that dinesh was prosecuted for our rulingcruz said rosie odonnell apparently committed the same offense five times five times when she broke the identical law that dinesh was prosecuted for says rosie odonnell apparently committed the same offense five times five times when she broke the identical law that dinesh (dsouza) was prosecuted for
wikileaks leak email	the revelation came out in 2016 in emails leaked by wikileaks linked to clinton campaign chairman john podesta who the todods invited to the party sarah palin who had her own emails leaked by wikileaks sang praises of the organization at an event in dc on thursday evening (cap was founded by john podesta who certainly cares a lot about russian attacks: when he was chair of hillary clintons campaign his emails were hacked by russian operatives and leaked by wikileaks through the final four weeks of the 2016 campaign)
child use tool	this rule could prevent children under 18 from using such tools as a power screwdriver a milking machine or something as simple as a wheelbarrow on the family farm another staff member filled gatorade bottles with liquid and glitter as a tool that children can use to become centered
wealthy american pay tax	says cnns wolf blitzer was wrong to say that the wealthiest americans they pay the most in taxes already 50 percent of americans dont even pay any federal income tax surveys are showing overwhelming support for raising taxes on top earners including a new politico/morning consult poll released monday that found 76 percent of registered voters believe the wealthiest americans should pay more in taxes
trump bail wall street main street	trump bailed out wall street but not main street
person enter united state	the attorney general has reported to me that more than 300 persons who entered the united states as refugees are currently the subjects of counterterrorism investigations by the federal bureau of investigation another thing that struck me was just how many persons illegally entering the united states have criminal records
democrat control senate	the scale of bidens plans now however hinge on whether democrats will narrowly control the senate or if all legislation must be negotiated with senate majority leader mitch mcconnell democrats have controlled the senate for 18 years of this period 77yearold joe biden thinks the democrats control the senate
flu kill american	the coronavirus now looks like it will soon claim more lives in the us than the influenza outbreaks of the 1950s and 60s that resulted in more than 100000 deaths — the worst pandemics in modern history behind only the 1918 flu that killed about 675000 americans the flu killed 80000 americans last year according to the cdc two weeks ago the cdc disclosed that more americans were killed by the flu last year than at any time in 40 years—more than 79000—and hospitalizations and illnesses hit record highs
president cut cdc budget	the president is cutting the cdc budget by 80% just last week the president defended his quest to cut the cdcs budget by insisting government can hire health experts whenever their services are needed
administration not-create policy separate family border	this administration did not create a policy of separating families at the border how is this not child abuse?this administration did not create a policy of separating families at the border nielsen said during the briefing later bristling when asked if the policy was intended to stir up outrage
jeff session justice department brought investigation popular republican congressman	two long running obama era investigations of two very popular republican congressmen were brought to a well publicized charge just ahead of the midterms by the jeff sessions justice department two long running obama era investigations of two very popular republican congressmen were brought to a well publicized charge just ahead of the midterms by the jeff sessions justice department trump wrote two long running obama era investigations of two very popular republican congressmen were brought to a well publicized charge just ahead of the midterms by the jeff sessions justice department trump tweeted
president trump sign executive order	just weeks after taking office president trump signed executive order 13788 buy american i this robust order immediately imposed greater oversight to stop unnecessary waivers to federal contractors the decision comes just a day after google ads attempted to demonetize the federalist over race based content and just a month after president trump signed an executive order with intentions to curb the cut back on the protections after alleged censorship of conservative news outlets under an executive order signed by president trump groups that resettle refugees must submit their plans including agreements with states and local governments by jan 21

Table D.19: All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); *cont.*

Narrative label	Sentences containing narrative label
president launch nuclear weapon	hillary clinton revealed a classified response window of four minutes for a us president to launch nuclear weapons legislators including house armed services committee chairman adam smith (d-wash) have already introduced bills to prevent presidents from acting solely on their own to launch nuclear weapons and to make it official policy that america will never initiate a nuclear war pelosi told her democratic colleagues during a call friday that she has gotten assurances there are safeguards in place in the event the president wants to launch a nuclear weapon according to sources on the call
asylum seeker travel mexico	over 500 members of a caravan of asylum seekers traveling through mexico to the us border at san diego were arrested in possession of suicide belts he said that the mexican delegation had moderated some of the us requests such as a safe third country agreement which would require asylum seekers traveling through mexico to make their case for american asylum in mexico
donald trump wear blackface	as ebrard noted in a news conference after the agreements announcement friday the mexican government did not accept the uss push for a safe third country agreement which would require asylum seekers traveling through mexico to make their case for american asylum in mexico donald trump was photographed wearing blackface can you even begin to imagine how the same people defending trudeau today would react if this was donald trump whod been caught wearing blackface
house democrat pass bill	a provision in the first bill passed in by house democrats would allow the billions of dollars of aid we send to other countries to be used for abortions republicans offered a package of roughly 1 trillion while the bill passed by house democrats was around 3 trillion 117k shares facebook twitter whatsapp pinterest reddit print mail flipboardsenate majority leader mitch mcconnell (r-ky) is complaining that democrats are obstructing coronavirus relief but house democrats passed the bill in may
people work hour	unemployment is low because people are working 60 70 80 hours a week and can barely feed their family productivity is very high people are working more hours than ever and incomes are stagnant the average number of hours people worked also edged up
vast number additional people storm border	but the problem is when you do that vast numbers of additional people storm the border
joe biden run president united state	sarah palin got more votes running for mayor of wasilla alaska than joe biden got running for president of the united states despite his lowkey campaign it remains the case that joe biden is running for president of the united states and that he had elected to take part in a televised event during which he was expected to take positions and address issues joe biden is of course running for president of the united states
president trump need magic wand	trump: president trump would need a magic wand to get to 4% gdp stated president obama claimed that former president obama once said president trump would need a magic wand to get to 4% gdp
pentagon fund medicare	president trump would need a magic wand to get to 4% gdp stated president obama medicare for all costs ~\$32t that means 66% of medicare for all could have been funded already by the pentagon that means 66% of medicare for all could have been funded already by the pentagon
president donald trump cut funding	medicare for all costs ~\$32t that means 66% of medicare for all could have been funded already by the pentagon says president donald trump cut funding from the justice department to investigate these domestic terrorist groups these white supremacist groups
man earns dollar	this follows claims that the world health organisation its public health body failed in its duty by slavishly following china which has led president donald trump to cut all us funding on average american women only earn 80 cents for every dollar a man earns for doing the same job in a job market where women make 80cents for every dollar a man earns female workers say they want leaders who care about inclusivity diversity and maternity leave among other employee benefits
co sponsor bill	for every dollar a man earns on average women earn 18 cents less or \$10194 less per year barack obama has written and cosponsored more bills in three years than hillary clinton has in seven kamala harris beto orourke cory booker and elizabeth warren cosponsored bill for abortions up to birth you do recognize that its in enshrined in the constitution the fourteenth amendment?i happen to be [cosponsoring] a bill that would say otherwise sanford stuttered
bill repeal obamacare	complaints about the lack of transparency behind the creation of a bill to repeal obamacare were hypocritical because the democrats were just as bad when they passed obamacare now interestingly as obamacare has come under scrutiny with a gopled congress and from president donald trump's administration the aarp has as part of a larger coalition went all in to try to protect obamacare--and the system that makes the group billions of dollars the washington post for instance in 2017 after the house passed a bill that would have repealed obamacare but protected preexisting coverage--and removed the aarps medigap loophole--reported that aarp was part of a coalition of medical and consumer groups during his first speech to a joint session of congress in february 2017 trump explained his goals for a bill to repeal obamacare
journalist news anchor spin news	cnn invested in an industrialized washing machine to help their journalists and news anchors spin the news before publication in order to aid the news station in preparing stories for consumption popular news media organization cnn purchased an industrialized washing machine to help its journalists and news anchors spin the news before publication the satire piece said
donald trump run president	then in early 2015 came another lifechanging email:brad pascare: it said donald trump is thinking about running for president donald trump has been disqualified from running for president due to his comments about muslims this was the worst deal in us and world history and the last straw and reason donald trump decided to run for president

Table D.20: All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); *cont.*

Narrative label	Sentences containing narrative label
business turn gay customer	senator ted cruz said that businesses should exercise their right to turn away gay customers buttigieg was mayor of south bend when the indiana governor signed a law in 2015 allowing businesses to turn away gay customers
donald trump drop presidential race	donald trump dropped out of the presidential race in august 2016 due to a brain tumor donald trump has dropped out of the 2016 presidential race and endorsed hillary clinton
democrat control congress	sen john thune (rsd) an establishment republican who once urged donald trump to drop out of the 2016 presidential race is slowwalking a trump nominee to the fcc whose confirmation is critical to tackling big tech censorship president joe biden and the democrats who now control congress could earn respect if not votes for telling coal country the truth that coal must be phased out of the national energy mix but that workers will not be left behind if the congress impeaches and removes trump and pence it will only be because the democrats control congress although president joe biden would rather monologue about infrastructure than talk about this and kamala harris will deny it a powerhungry biden administration in cahoots with a democratcontrolled congress is working hard to enact sweeping gun control
people cast vote	instead the data seems to suggest that voter purges could simply make it tougher for many people to cast their vote at all vote for hillary and be a part of historyaccording to the complaint more than 4900 people 'cast their vote via text message
iran release american hostage	he also alleged that there were some outofstate voters in georgia and people who had cast votes twice there says bill nelson voted against requiring iran to recognize israel and release american hostages this request includes but is not limited to all related records of communications between any official employee or representative of any of the following entities:the department of the treasurythe islamic republic of iranthe executive office of the white housethe department of defensethe department of justicethe department of statein january 2016 the funds sent in stacks of foreign currency on wooden pallets on an unmarked cargo plane were delivered to tehran about the same time iran released five american hostages
lobbyist run campaign	i dont have lobbyists running my campaign trump vowed to 'drain the swamp but lobbyists are helping run his campaign
century law open floodgate special interest include foreign corporation spend limit election	barack obama: with all due deference to separation of powers last week the supreme court reversed a century of law that i believe will open the floodgates for special interests — including foreign corporations — to spend without limit in our elections last week the supreme court reversed a century of law that i believe will open the floodgates for special interests – including foreign corporations – to spend without limit in our elections in his 2010 state of the union address president barack obama criticized the citizens united ruling in front of members of the supreme court saying it had reversed a century of law that i believe will open the floodgates for special interests — including foreign corporations — to spend without limit in our elections
email damage hillary clinton campaign	mr stone told mr trump that he had just gotten off the phone with [wikileaks founder] julian assange and that mr assange told mr stone that within a couple of days there would be a massive dump of emails that would damage hillary clintons campaign mr stone told mr trump that he had just gotten off the phone with julian assange and that mrasange told mr stone that within a couple of days there would be a massive dump of emails that would damage hillary clintons campaign stone who reportedly keeps in touch with trump boasted last year that he had a back channel to julian assange the wikileaks founder and predicted the release by wikileaks of emails that damaged hillary clintons campaign
american buy health insurance	says president donald trump's proposal to allow americans to buy health insurance across state lines would make it the same as the way you buy car insurance today health reform will give every american the same opportunity to buy health insurance the way members of congress do
family lose home	the first is a public option in health care allowing americans to buy health insurance from the government she talks of her familys losing its home after her mother couldnt pay the bills after her army service charges against the teenager were later dropped but that didnt spare the family from losing their home dewine took \$19 million from big banks supported legislation that helped bernie madoff make millions and protected predatory lenders while families lost their homes
law separate child	bill clinton passed a law in 1996 that separated children from adults illegally entering america comyn was one of a number of republicans who said the choice comes down to a decision between enforcing laws separating children from parents who are charged with crimes and ignoring those laws he said at the time: if you are smuggling a child then we will prosecute you and that child will be separated from you as required by law
member leave office	statement: in 13 of the 15 cities all but phoenix and san diego at least one member has faced an indictment or investigation for a serious offense in the past decade while in office or soon after leaving office according to a feb 2 report by the pew charitable trusts meanwhile across the pond: plenty of former members of congress register as lobbyists shortly after leaving office
american lose job	more than 15000 americans were losing their jobs each day in june 2009 as the us struggled to climb out of a painful recession following its worst financial crisis in decades she insists on playing partisan politics and putting the democrat partys farleft agenda ahead of peoples lives while americans are dying and losing their jobs of course the democrats ran one ad after another showing the americans who lost their jobs to make mitt filthy rich as their factories were shuttered and shipped to china
fbi spy campaign	having spent two years trying to discredit muellers work as a witchhunt and hoax he stepped up demands for an investigation into its origins and pushed the claim that the fbi spied on his 2016 campaign claimed former intelligence director james clapper said trump should be happy that the fbi was spying on his campaign then when trump said an even bigger scandal is the way the fbi spied on his campaign stahl claims theres no real evidence of that and says this is '60 minutes and we cant put on things that we cant verify
healthy people wear mask	a patient discharge document proves that the who and cdc do not recommend that healthy people wear masks people who are sick should wear masks said adams but acknowledged that if healthy people feel better by wearing a mask by all means wear it but they should not touch their faces it was a reversal for the agency that said at the beginning of the outbreak healthy people did not need to wear masks because it would not protect them from contracting the disease
federal constitutional amendment ban gay marriage	says rick perry flipfopped on gay marriage by first saying it was a states rights issue and then saying he was in favor of a federal constitutional amendment banning gay marriage he also supported a federal constitutional amendment that would ban gay marriage

Table D.21: All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); *cont.*

Narrative label	Sentences containing narrative label
fbi director james comey leak classify information	says former fbi director james comey admitted to leaking classified information the former mayor also called for former fbi director james comey to be prosecuted for leaking classified information at one point calling him a disgraceful liar
mexico afford build wall	says mexico can afford to build a wall because the countrys trade deficit with america is billions of dollars back on the campaign trail trump claimed that mexico can afford to build a wall because mexicos trade surplus with america is billions of dollars
major hurricane strike land	says 11 straight years of no major hurricanes striking land in the united states bores a hole right through the whole climate change argument 11 straight years of no major hurricanes striking land in the united states bores a hole right through the whole climate change argument
president deems statute unconstitutional	when the president deems the statute unconstitutional is contrary to 200 years of supreme court precedent under the constitution kavanaugh wrote the president may decline to enforce a statute that regulates private individuals when the president deems the statute unconstitutional even if a court has held or would hold the statute constitutional in a footnote kavanaugh wrote that the president may decline to enforce a statute that regulates private individuals when the president deems the statute unconstitutional
iran release prisoner	the obama administration secretly bribed iran to release four prisoners in january 2016 this claim from cotton refers to a \$400 million payment the obama administration made to iran in january 2016 on the day iran released several us prisoners and implemented the (since disbanded) nuclear deal
people game system	and then we saw even more sophisticated attacks around that where people found out ways to game the systems in order to get more visibility and to get their message higher than someone else under the affordable care act some people were gaming the system by paying for nine months of insurance get 12 the other three for free people would do whatever they could to game the system — a system that was designed to keep people out instead of letting them in
republican replace obamacare	says the republican plan to replace obamacare absolutely does not eliminate protections for preexisting conditions vice president pence said thursday that republicans plan to give another try to repealing and replacing obamacare if they do well enough in novembers elections namely jordan is adamant that republicans need to get back to accomplishing what they promised voters during the 2016 election cycle like dealing with immigration border security repealing and replacing obamacare and stopping outofcontrol spending
drug induce abortion	in answer to a question from sen ted cruz (rtx) kavanaugh said they said filling out the form would make them complicit in the provision of the abortioninducing drugs that they were as a religious matter objecting to a georgia religious freedom bill would help catholic institutions that under obamacare could be forced to provide abortioninducing drugs
police open fire	the washington post politifact and others refuted their claims that kavanaugh called birth control abortioninducing drugs but the proabortion politicians continue to push the lie thats exactly where i was standing when the police opened fire both brooks and the officer immediately behind him are seen holding police tasers with illuminationbrooks (far right) appears to turn and point the stolen taser as one officer takes cover by the red car and the second officer pursues from the far left of the imagebrooks drops the the ground out of sight behind the silver sedan as police open fire on himpreviously released bystander video shows brooks in a scuffle as officers tried to take him into custody in a parking lot of a wendys on university avenue at around 10:30pm on friday police opened fire on a store mannequin after it failed to follow their orders
obama administration try iran secret access financial system united state	the obama administration is now accused of trying to give iran secret access to the financial system of the united states
donald trump grope woman	leaked dnc emails from may 2016 reveal a plot to contrive fake stories about donald trump groping women she celebrated nordstroms decision as a milestone for the campaign which began in october after the washington post obtained a video from 2005 that showed donald trump bragging about groping women during a taping of access hollywood
donald trump encourage supporter	donald trump encouraged supporters to engage in a nazistyle salute at a campaign rally donald trump routinely uses scripted violence and eliminationist rhetoric to encourage his supporters to target his enemies trump supporters are making it more likely that stayathome orders last longerthe protests that donald trump has encouraged his supporters to participate in will only make it more likely that the stayathome orders will last longer
china build additional coal plant ingredient cause cancer	china will be allowed to build hundreds of additional coal plants snacks marketed to children are heavily laden with ingredients that cause cancer last year the fda issued an alert over a cancercausing ingredient used in the blood pressure medication valsartan made by chinese company zhejiang huahai resulting in a recall of affected drugs pepsi admitted that their soda contains cancercausing ingredients
hanoi hilton songbird collaborate enemy	says mccain was a hanoi hilton songbird who collaborated with the enemy
noise windmill cause cancer	it says that when he was a pow mccain was a hanoi hilton songbird who collaborated with the enemy the noise (from windmills) causes cancer one of the most bonkers they say items: his 2019 declaration that they say the noise from windmills causes cancer
group support president donald trump	a group supporting president donald trump is making robocalls threatening people who criticize him online the group supporting president donald trump were carrying make america great again signs and american flags as they marched to beverly gardens park
special interest alien try come country southern border	cbp has stopped over 3000 what we call specialinterest aliens trying to come into the country on the southern border in context wallace showed a clip of homeland security secretary kirstjen nielsen claiming that customs and border patrol has stopped over 3000 what we call special interest aliens trying to come into the country in the southern border
black kill black	10856 blacks were killed by other blacks in 503 days in us that prepackaged answer ignores that just as more than 80% of blacks are killed by blacks so too are more than 80% of whites killed by whites yet how many blacks have taken to the streets to kill other blacks?

Table D.22: All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); *cont.*

Narrative label	Sentences containing narrative label
queen elizabeth ii offer restore british rule united state america	queen elizabeth ii has offered to restore british rule over the united states of america london (the borowitz report)—in an unexpected televised address on saturday queen elizabeth ii offered to restore british rule over the united states of america though relatively benign at first (al goes diet is making him stupid) the trump/pecker relationship has metastasized: in effect the enquirer became an enforcement arm of the trump presidential campaign and presidency as the us attorney in the southern district of new york laid out in its case against michael cohen who has pleaded guilty
michael cohen plead guilty	a sixth investigative avenue was opened this spring by federal prosecutors in the southern district of new york where michael cohen pleaded guilty to violating campaign finance laws by paying hush money to stormy daniels and others—which he says occurred at trumps instruction the newest bombshell is the story that michael cohen is now pleading guilty to lying about when the trump tower in moscow deal actually fell through
hillary clinton abolish second amendment	says hillary clinton wants to abolish the second amendment says hillary clinton wants to essentially abolish the second amendment trump regularly told crowds on the campaign trail that hillary clinton would abolish the second amendment and collect their guns if she was elected
terrorist kill american	dems hate trump more than they hate terrorists who kill americans click here to get the fox news appthink about what she said when 3000 americans were killed by terrorists on 9/11 ratcliffe said says former president barack obama pardoned a traitor who gave us enemies state secrets he also pardoned a terrorist who killed americans
million lose health care coverage	millions could lose their health care coverage and be forced into a government pool under the new health care law efforts to contain the pandemic also caused a devastating recession which has resulted in millions losing their healthcare coverage the stakes are extraordinarily high for the american people as millions could lose their health care coverage feinstein said
health insurance preexist condition	david perdue says his policy is that health insurance should always cover preexisting conditions that for the first time since obamacare passed they will allow health insurance companies to once again start refusing to cover people with preexisting conditions
missile knock missile air	we have missiles that can knock out a missile in the air 97 percent of the time and if you send two of them its going to get knocked down oct 11 we have missiles that can knock out a missile in the air we have missiles that can knock out a missile in the air
democrat secure border	democrats have always been for smart and effective ways to secure our border schumer said democrats oppose any effort to secure our border democrats have always been for smart and effective ways to secure our border
republican win wisconsin	is trump the first republican to win wisconsin since 1952? president trump won the state by fewer than 23000 votes in 2016 becoming the first republican to win wisconsin since 1984 after democratic nominee hillary clinton infamously failed to campaign in the state
company ship job	says ron johnson helped companies ship jobs overseas says the koch brothers want to protect tax cuts for companies that ship our jobs overseas he said hed force companies to keep jobs in america and that there would be consequences for companies that shipped jobs abroad
oft use phrase	actress betty white uttered a witty rejoinder to the oftused phrase grow some balls that oftused phrase among birth workers reflects the protective value of black sisterhood; the comfort of feeling valued and understood
president trump marshal resource federal government	president trump marshalled the full resources of our federal government from the outset in pences telling president trump marshalled the full resources of our federal government from the outset
people play game	you want people to stop hiding the truth and playing games with you and if people want to play that game we could look into why an evangelical christian like mike pence wants to be on a ticket with the president caught with a porn star or how he feels about the immigration policy that he called ‘unconstitutional before he decided to team up with donald trump we dont show people playing these games
woman file lawsuit	a woman filed a lawsuit against samsung after her cellphone became stuck in her vagina advertisementadvertisementadvertisementless than two weeks after those women filed the lawsuit cone faced a graver accusation: a woman in her early 20s told police in philadelphia that hed raped her women who said they had been harassed filed only two of the lawsuits

Table D.23: All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); *cont.*

Narrative label	Sentences containing narrative label
government run healthcare	says donald trump promised to take care of everybody (through governmentrun healthcare and) the governments going to pay for it club for growth says donald trump promised to take care of everybody (through governmentrun healthcare and) the governments going to pay for it but remember where the democrats are now on healthcare where more than 130 democrats have cosponsored—they want a governmentrun healthcare that for any of your listeners right now who are on medicare they are going to bankrupt medicare
american abide law	the united states of america should be a sanctuary for lawabiding americans not criminal aliens! this has been a year unlike any other where lawabiding americans are seeing for themselves the need to be able to protect themselves and their family olivia stated pictwittercom/spud5n3ixg — oc sheriff don barnes (@ocsheriffbarnes) april 29 2020welcome to 2020 where sex offenders and child molesters are being released from prisons while lawabiding americans are being arrested harassed and fined for peacefully protesting and going to church
people work health care	theres 400 more people working in health care today in new hampshire than when we passed the budget but the remaining 27 percent said they would probably or definitely never take it a proportion that rose to 33 percent among black adults 33 percent among essential workers and 29 percent among people working in health care when betz and victoroff started the group in 2015 they knew they wanted to include gun dealers as well as people who work in health care
child care tax credit	republicans dont think its a good idea to make the child care tax credit stronger rep seth moulton (dma)moulton told vox in a statement that he supports a recent proposal by the advocacy group the childrens defense fund which would expand the child tax credit and the child care tax credit and make them fully refundable as well as making child care subsidies available to families making up to 150 percent of the poverty line
person commit suicide	a person can be seen committing suicide by jumping from a bridge in the background of a scene in the movie the basketball diaries here are some recommendations from professionals and endorsed by the cdc:dont glorify the act or the persons who commit suicide
immigrant cross southern border	a form of tuberculosis dubbed mexican blood flu is being brought into the us by immigrants crossing the southern border breaking: judge bars trump administration from refusing asylum to immigrants who cross southern border illegally a federal judge in san francisco on monday barred the trump administration from refusing asylum to immigrants who cross the southern border illegally likely prompting a legal challenge from the white house
illegal immigrant receive driver license	and in ohio dewine allowed illegal immigrants to receive drivers licenses senator marsha blackburn (r tenn) announced on a conference call wednesday that she was proposing a bill to block grant funding from the department of justice for local law enforcement in states with sanctuary policies that allow illegal immigrants to receive drivers licenses new york makes it felony for officials to share illegal immigrant driver data with icewkbw first reported that the law which went into effect in january and primarily allows illegal immigrants to receive drivers licenses was amended as part of the states 2020 budget and signed by democratic gov andrew cuomo in april
people speak spanish	reading the language is necessary to be a professional to understand and converse with people who speak spanish in 5 people in the us speak spanish? the government labels such migrants special interest aliens so they can be flagged to ostensibly undergo enhanced security screenings such as threat assessment interviews to which spanishspeaking people are not subject
donald trump won election	donald trump won the 2020 election it wasnt that long ago that obama deep state operatives led by us deputy attorney general sally yates accused general michael flynn of violating the logan act for making calls to foreign dignitaries including the russian ambassador after donald trump won the election in 2016 he also criticized critical race theory and the handling of the covid19 pandemic by the cdcand of course there was the obligatory end to gaetzs monologue: donald trump won the election!
people support obamacare	the numbers of people that support obamacare and like it have been steadily dropping many people grudgingly supported obamacare because they were convinced uninsured people were imposing backbreaking costs upon the health care system by flooding emergency rooms and obliging hospitals to treat them
chicago experience high employment	chicago is experiencing right now the highest employment since 1950 current issue view our current issuewhen it comes to jobs emanuel likes to repeat the claim that chicago is experiencing the highest employment since 1950

Table D.24: All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); *cont.*

Narrative label	Sentences containing narrative label
republican pass bill	says wisconsin women facing pay discrimination cant do something about it under bill passed by republicans republicans passed this bill partially in the hope that democrats would be baited into running against tax cuts and for raising taxes and pelosi seems to be obliging them
illegals commit crime	@kenpaxtonx: weve had over 600000 crimes committed by illegals since 2011 says texas has had over 600000 crimes committed by illegals since 2011 previous administrations have deliberately kept americans in the dark about illegal immigrant crimesmost states and our federal government have kept information and statistics about illegal immigration crimes committed by illegals and the costs borne by you the us payer out of public view
supreme court reverse century law believe open floodgate special interest include foreign corporation spend limit election	barack obama: with all due deference to separation of powers last week the supreme court reversed a century of law that i believe will open the floodgates for special interests — including foreign corporations — to spend without limit in our elections last week the supreme court reversed a century of law that i believe will open the floodgates for special interests — including foreign corporations — to spend without limit in our elections in 2010 shortly after the controversial citizens united decision thenpresident barack obama rebuked the members of the supreme court seated before him in the state of the union:with all due deference to separation of powers last week the supreme court reversed a century of law that i believe will open the floodgates for special interests — including foreign corporations — to spend without limit in our elections
wife run office	the man that was investigating her from the fbi his wife runs for office and they give her more than \$675000 to run internal fbi documents stated that mccabe didnt oversee the clinton investigation while his wife was running for office and that he didnt have a conflict of interest
president reduce deficit	i know the president certainly would like to reduce the deficit and its one of the reasons that his budget — this budget reduced the deficit by \$3 trillion which was one of the largest in history the president aims to reduce the deficit by 44 trillion over a decade by way of about 2 trillionin spending restraint through nondefense programs and about \$2 trillion from military accounts plus hundreds of billions of dollars in avoided interest payments
secretary state secpompeo study south africa land farm seizure expropriation large scale kill farmer	i have asked secretary of state @secpompeo to closely study the south africa land and farm seizures and expropriations and the large scale killing of farmers this action only proves president trump and his department of justice would rather make political points than defend our health care law that protects the 133 million americans with preexisting conditions has provided tens of millions more americans with health insurance coverage and has lowered prescription drug costs for millions of seniors schumer said in a statement
pre exist condition	sen collins owes it to her constituents to hear them out and explain why judge kavanaughs demonstrated hostility to roe v wade opposition to covering preexisting conditions and repeated lies is not disqualifying current data shows that if you have no preexisting conditions your fatality rate if you contract the coronavirus is 9% and that includes the elderly
republican repeal affordable care act	but since 2016 republicans have tried and failed to repeal the affordable care act and the law has gained popularity but adams watched with disgust as she and other republicans attempted to repeal the affordable care act in 2017 and later as the president proposed limiting medicare spending and cutting social security disability benefits despite promising to protect both programs as republicans try to repeal the affordable care act they should be reminded every day that 36000 people will die yearly as a result
help save life	new stroke indicator can help save lives two of them helped saved my life after i was hit by that car in 2005
medicare save trillion	let me thank the koch brothers of all people for sponsoring a study that shows that medicare for all would save the american people \$2 trillion over a 10year period democrats seize on cherrypicked claim that ‘medicareforall would save \$2 trillion
china paid tariff	authorized \$28 billion in aid for farmers who have been subjected to unfair trade practices — fully funded by the tariffs paid by china thats another way in which the tariffs are paid not by china but by us mr trump tariffs are paid for mostly by china by the way not by us
mexican enter country	in august 2014 michele bachmann told fox news that banning mexican food in the united states would stop mexicans from entering the country without documentation we think the decline in the number of unauthorized immigrants was almost entirely due to fewer mexicans entering the country without authorization

Table D.25: All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); *cont.*

Narrative label	Sentences containing narrative label
democrat commit fraud	video depicts democrats committing fraud during the 2016 primary another person who submitted an affidavit had posted on facebook that democrats had planned to commit fraud kenny noted writing that his predilection to believe fraud was occurring undermines his credibility as a witness pounding his claim — which is now as familiar as it is demonstrably false — that he lost the election due to fraud committed by democrats trump struck a match and watched the fire spread
donald trump release tax return	says donald trump hasnt broken his promise to release his tax returns this revelation is the latest reminder that every single day donald trump is breaking his promise to the american people to release his tax returns and be transparent about for whom he really works biden campaign spokesman andrew bates said
mother meet doctor	with a lateterm abortion the mother meets with the doctor the mother meets with the doctor the baby is born; the mother meets with the doctor he told the crowd
noise cause cancer	and they say the noise causes cancer and they say the noise causes cancer he said heres video of trump talking about wind mills at the nrcc dinner tonight: they say the noise causes cancer
woman perform abortion	a woman performed her own abortion with a uabort it™ attachment for her ar15 in 12 years she estimates theyve helped 70000 women perform their own abortions she and another member of her group carol downey began touting the device as a way women could perform abortions on each other safely
bill clinton sign nafta	nafta signed by bill clinton has been a disaster for pennsylvania but trump threw all that threeyardsandacloudofdust work into the dustbin of history so the nafta in place today is the same nafta envisioned by ronald reagan negotiated by george hw bush and signed into law by bill clinton
republican control state legislature	the democratic governor of the state tony evers asked the republicancontrolled state legislature to move back the date of the election like so many other states have done amid concerns over covid19 gov scott walker and the republicancontrolled state legislature have added over \$140 million in new special interest spending mark levin a rightwing radio host with a penchant for hysteria urged republicancontrolled state legislatures to ignore the results of their state elections and send electors who will vote for trump in the electoral college
company avoid tax	facebook post says president barack obama signed the medical appliance tax bill that forced companies to outsource manufacturing of masks gowns gloves and ventilaors [sic] to china europe and russia to avoid the tax in december the uk campaign group fair tax mark put amazon at the top of a list of companies that it said were aggressively avoiding tax saying it had paid just 34bn in tax on revenues of 9605bn and profits of \$268bn sanders team however is expanding on the idea lowering the ratio companies would have to meet to avoid the tax and extending it to both public and privately held companies with an annual revenue of more than \$100 million
federal government rent donald trump golf cart protect donald trump round	the federal government has spent over \$150000 just the secret service alone for the privilege of renting donald trump golf carts to protect donald trump on his rounds
america defeat nazi	where were they in world war ii when america defeated the nazis democratic socialist congressional candidate alexandria ocasio-cortez wants to address global warming the same way america defeated the nazis during world war ii

Table D.26: All high-dimensional narratives automatically extracted from false claims texts and used to test our hypothesis ($n = 318$, see Table 5.3); *cont.*

Bibliography

- Alan I Abramowitz. 1991. Incumbency, campaign spending, and the decline of competition in us house elections. *The Journal of Politics*, 53(1):34–56.
- Hirofugu Akaike. 1974. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723.
- Michael Olusegun Akinwande, Hussaini Garba Dikko, Agboola Samson, et al. 2015. Variance inflation factor: as a condition for the inclusion of suppressor variable (s) in regression analysis. *Open journal of statistics*, 5(07):754.
- Nikolaos Aletras and Mark Stevenson. 2013. **Evaluating topic coherence using distributional semantics**. In *International Conference on Computational Semantics (IWCS)*. Association for Computational Linguistics.
- Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2):211–36.
- Hunt Allcott, Matthew Gentzkow, and Chuan Yu. 2019. Trends in the diffusion of misinformation on social media. *Research & Politics*, 6(2):2053168019848554.
- Jennifer Allen, Baird Howland, Markus Mobius, David Rothschild, and Duncan J Watts. 2020. Evaluating the fake news problem at the scale of the information ecosystem. *Science advances*, 6(14):eaay3539.
- Daniel Allington, Sarah Brouillette, and David Golumbia. 2016. **Neoliberal tools (and archives): A political history of digital humanities**. In *LA Review of Books*.
- Areej Alokaili, Nikolaos Aletras, and Mark Stevenson. 2019. **Re-ranking words to improve interpretability of automatically generated topics**. In *Proceedings of the 13th International Conference on Computational Semantics - Long Papers*, pages 43–54, Gothenburg, Sweden. Association for Computational Linguistics.
- Sacha Altay, Anne-Sophie Hacquin, and Hugo Mercier. 2022a. Why do so few people share fake news? it hurts their reputation. *New Media & Society*, 24(6):1303–1324.

- Sacha Altay, Rasmus Kleis Nielsen, and Richard Fletcher. 2022b. Quantifying the “infodemic”: People turned to trustworthy news outlets during the 2020 coronavirus pandemic. *Journal of Quantitative Description: Digital Media*, 2.
- Hadi Amiri, Philip Resnik, Jordan Boyd-Graber, and Hal Daumé III. 2016. Learning text pair similarity with context-sensitive autoencoders. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1882–1892.
- Aris Anagnostopoulos, Ravi Kumar, and Mohammad Mahdian. 2008. Influence and correlation in social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 7–15. ACM.
- Stephen Ansolabehere, John M De Figueiredo, and James M Snyder Jr. 2003. Why is there so little money in us politics? *Journal of Economic perspectives*, 17(1):105–130.
- Stephen Ansolabehere and Philip Edward Jones. 2010. Constituents’ responses to congressional roll-call voting. *American Journal of Political Science*, 54(3):583–597.
- Stephen Ansolabehere, James M Snyder Jr, and Charles Stewart III. 2001. The effects of party and preferences on congressional roll-call voting. *Legislative Studies Quarterly*, pages 533–572.
- Sinan Aral and Dean Eckles. 2019. Protecting elections from social media manipulation. *Science*, 365(6456):858–861.
- Daniel Argyle, Lisa P Argyle, Vlad Eidelman, and Philip Resnik. 2021. Debate reaction ideal points: Political ideology measurement using real-time reaction data. *Statistics, Politics and Policy*, 12(1):5–28.
- Daniel Argyle, Lisa P. Argyle, Vladimir Eidelman, and Philip Resnik. 2019. Ideal point estimation using real-time presidential debate reactions.
- Douglas Arnold. 1990. R. 1990 the logic of congressional action.
- Elliott Ash, Germain Gauthier, and Philine Widmer. 2021. Text semantics capture political and economic narratives. *Center for Law & Economics Working Paper Series*, 2021(11):2108–01720.
- Isabelle Augenstein, Christina Lioma, Dongsheng Wang, Lucas Chaves Lima, Casper Hansen, Christian Hansen, and Jakob Grue Simonsen. 2019. Multifc: A real-world multi-domain dataset for evidence-based fact checking of claims. pages 4685–4697.
- Peter Bachrach and Morton S Baratz. 1962. Two faces of power. *American Political Science Review*, 56(4):947–952.

- Peter Bachrach and Morton S Baratz. 1963. Decisions and nondecisions: An analytical framework. *American political science review*, 57(3):632–642.
- Joseph Bafumi and Michael C Herron. 2010. Leapfrog representation and extremism: A study of american voters and their members in congress. *American Political Science Review*, 104(3):519–542.
- Anne E Baker. 2018. Are federal pacs obsolete? *Interest Groups & Advocacy*, 7(2):105–125.
- Robert E Baldwin and Christopher S Magee. 2000. Is trade policy for sale? congressional voting on recent trade bills. *Public Choice*, 105(1):79–101.
- Andrew O Ballard, Ryan DeTamble, Spencer Dorsey, Michael Heseltine, and Marcus Johnson. 2023. Dynamics of polarizing rhetoric in congressional tweets. *Legislative Studies Quarterly*, 48(1):105–144.
- Inger Baller. 2017. Specialists, party members, or national representatives: Patterns in co-sponsorship of amendments in the european parliament. *European Union Politics*, 18(3):469–490.
- Antoine Banks, Ernesto Calvo, David Karol, and Shibley Telhami. 2021. # polarizedfeeds: Three experiments on polarization, framing, and social media. *The International Journal of Press/Politics*, 26(3):609–634.
- João Pedro Baptista and Anabela Gradim. 2020. Understanding fake news consumption: A review. *Social Sciences*, 9(10):185.
- Michael Barber. 2016. Donation motivations: Testing theories of access and ideology. *Political Research Quarterly*, 69(1):148–159.
- Pablo Barberá. 2015. Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political analysis*, 23(1):76–91.
- Pablo Barberá. 2020. Social media, echo chambers, and political polarization. *Social media and democracy: The state of the field, prospects for reform*, 34.
- Pablo Barberá, Andreu Casas, Jonathan Nagler, Patrick J Egan, Richard Bonneau, John T Jost, and Joshua A Tucker. 2019. Who leads? who follows? measuring issue attention and agenda setting by legislators and the mass public using social media data. *American Political Science Review*, 113(4):883–901.
- Chris Barker and Dariusz Galasinski. 2001. *Cultural studies and discourse analysis: A dialogue on language and identity*. Sage.

- David A Bateman and John Lapinski. 2016. Ideal points and american political development: Beyond dw-nominate. *Studies in American Political Development*, 30(2):147–171.
- James Coleman Battista, Michael Peress, and Jesse Richman. 2021. Estimating the locations of voters, politicians, policy outcomes, and status quos on a common scale. *Political Science Research and Methods*, pages 1–17.
- Raymond Augustine Bauer, Ithiel de Sola Pool, and Lewis Anthony Dexter. 1963. *American business and public policy: the politics of foreign trade*. New York: Atherton Press.
- Frank R Baumgartner, Jeffrey M Berry, Marie Hojnacki, Beth L Leech, and David C Kimball. 2009. *Lobbying and policy change: Who wins, who loses, and why*. University of Chicago Press.
- Frank R Baumgartner, Suzanna L De Boef, and Amber E Boydstun. 2008. *The decline of the death penalty and the discovery of innocence*. Cambridge University Press.
- Frank R Baumgartner and Bryan D Jones. 1993. *Agendas and Instability in American Politics*. University of Chicago Press.
- Frank R Baumgartner and Beth L Leech. 2001. Interest niches and policy bandwagons: Patterns of interest group involvement in national politics. *The Journal of Politics*, 63(4):1191–1213.
- Kathleen Bawn, Martin Cohen, David Karol, Seth Masket, Hans Noel, and John Zaller. 2012. A theory of political parties: Groups, policy demands and nominations in american politics. *Perspectives on Politics*, 10(3):571–597.
- Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 610–623.
- W Lance Bennett and Steven Livingston. 2018. The disinformation order: Disruptive communication and the decline of democratic institutions. *European journal of communication*, 33(2):122–139.
- Kenneth Benoit and Michael Laver. 2012. The dimensionality of political space: Epistemological and methodological considerations. *European Union Politics*, 13(2):194–218.
- Sabine Bergler. 1992. *Evidential analysis of reported speech*. Brandeis University.
- Sabine Bergler. 2006. Conveying attitude with reported speech. In *Computing attitude and affect in text: Theory and applications*, pages 11–22. Springer.

- William D Berry, Richard C Fording, Evan J Ringquist, Russell L Hanson, and Carl E Klarner. 2010. Measuring citizen and government ideology in the us states: A re-appraisal. *State Politics & Policy Quarterly*, 10(2):117–135.
- Anthony M Bertelli and Christian R Grose. 2011. The lengthened shadow of another institution? ideal point estimates for the executive branch and congress. *American Journal of Political Science*, 55(4):767–781.
- John C Besley and Matthew Nisbet. 2013. How scientists view the public, the media and the political process. *Public understanding of science*, 22(6):644–659.
- Shraey Bhatia, Jey Han Lau, and Timothy Baldwin. 2017. **An automatic approach for document-level topic model evaluation**. In *Conference on Computational Natural Language Learning*, Vancouver, Canada. Association for Computational Linguistics.
- Sima Bhowmik, Md Main Uddin Rony, Md Mahfuzul Haque, Kristen Alley Swain, and Naeemul Hassan. 2019. Examining the role of clickbait headlines to engage readers with reliable health-related information. *arXiv preprint arXiv:1911.11214*.
- Federico Bianchi, Silvia Terragni, and Dirk Hovy. 2021. **Pre-training is a hot topic: Contextualized document embeddings improve topic coherence**. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 759–766, Online. Association for Computational Linguistics.
- Edward Loper Bird. 2009. Steven and ewan klein. 2009. natural language processing with python.
- Benjamin G Bishin. 2003. Independently validating ideology measures: A look at nominate and adjusted ada scores. *American Politics Research*, 31(4):404–425.
- Prakhar Biyani, Kostas Tsioutsouliklis, and John Blackmer. 2016. "8 amazing secrets for getting more clicks": Detecting clickbaits in news streams using article informality. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30.
- David M Blei. 2012. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84.
- David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.
- Jonas Nygaard Blom and Kenneth Reinecke Hansen. 2015. Click bait: Forward-reference as lure in online news headlines. *Journal of Pragmatics*, 76:87–100.

- Rachel Blum, Lindsey Cormack, and Kelsey Shoub. 2022. Conditional congressional communication: how elite speech varies across medium. *Political Science Research and Methods*, pages 1–8.
- Adam Bonica. 2013. Ideology and interests in the political marketplace. *American Journal of Political Science*, 57(2):294–311.
- Adam Bonica. 2018. Inferring roll-call scores from campaign contributions using supervised machine learning. *American Journal of Political Science*, 62(4):830–848.
- Adam Bonica. 2019. Database on ideology, money in politics, and elections (dime).
- Adam Bonica, Gary W Cox, et al. 2018. Ideological extremists in the us congress: Out of step but still in office. *Quarterly Journal of Political Science*, 13(2):207–236.
- Adam Bonica, Howard Rosenthal, and David J Rothman. 2014. The political polarization of physicians in the united states: an analysis of campaign contributions to federal elections, 1991 through 2012. *JAMA Internal Medicine*, 174(8):1308–1317.
- Adam Bonica and Maya Sen. 2021. Estimating judicial ideology. *Journal of Economic Perspectives*, 35(1):97–118.
- Samuel R. Bowman, Luke Vilnis, Oriol Vinyals, Andrew Dai, Rafal Jozefowicz, and Samy Bengio. 2016. **Generating sentences from a continuous space**. In *Conference on Computational Natural Language Learning*. Association for Computational Linguistics.
- Janet M Box-Steffensmeier, Peter M Radcliffe, and Brandon L Bartels. 2005. The incidence and timing of pac contributions to incumbent us house members, 1993–94. *Legislative Studies Quarterly*, 30(4):549–579.
- Jordan Boyd-Graber, Yuening Hu, and David Mimno. 2017. *Applications of Topic Models*. NOW Publishers.
- Lia Bozarth and Ceren Budak. 2021. An analysis of the partnership between retailers and low-credibility news publishers. *Journal of Quantitative Description: Digital Media*, 1.
- Lia Bozarth, Aparajita Saraf, and Ceren Budak. 2020. Higher ground? how groundtruth labeling impacts our understanding of fake news about the 2016 us presidential nominees. 14:48–59.
- Stephen G Bronars and John R Lott, Jr. 1997. Do campaign donations alter how a politician votes? or, do donors support candidates who value the same things that they do? *The Journal of Law and Economics*, 40(2):317–350.

- David E Broockman and Daniel M Butler. 2017. The causal effects of elite position-taking on voter attitudes: Field experiments with elite communication. *American Journal of Political Science*, 61(1):208–221.
- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.
- Marc Brysbaert, Michaël Stevens, Paweł Mandera, and Emmanuel Keuleers. 2016. **How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant’s age.** In *Frontiers in Psychology*.
- Ceren Budak. 2019. What happened? the spread of fake news publisher content during the 2016 us presidential election. pages 139–150.
- Barry C Burden. 2007. *Personal roots of representation*. Princeton University Press.
- Joanna M Burkhardt. 2017a. *Combating fake news in the digital age*, volume 53. American Library Association Chicago, IL.
- Joanna M Burkhardt. 2017b. History of fake news. *Library Technology Reports*, 53(8):5–9.
- Sophie Burkhardt and Stefan Kramer. 2019. **Decoupling Sparsity and Smoothness in the Dirichlet Variational Autoencoder Topic Model.** In *Journal of Machine Learning Research*.
- Kenneth P Burnham and David R Anderson. 2004. Multimodel inference: understanding aic and bic in model selection. *Sociological methods & research*, 33(2):261–304.
- Christine Cahill and Andrey Tomashevskiy. 2019. Private donations and policy ambiguity. *Comparative Political Studies*, 52(5):754–776.
- Brandice Canes-Wrone, David W Brady, and John F Cogan. 2002. Out of step, out of office: Electoral accountability and house members’ voting. *American Political Science Review*, 96(1):127–140.
- Dallas Card. 2019. *Accelerating Text-as-Data Research in Computational Social Science*. Ph.D. thesis, University of Washington.
- Dallas Card, Serina Chang, Chris Becker, Julia Mendelsohn, Rob Voigt, Leah Boustan, Ran Abramitzky, and Dan Jurafsky. 2022. Computational analysis

- of 140 years of us political speeches reveals more positive but increasingly polarized framing of immigration. *Proceedings of the National Academy of Sciences*, 119(31):e2120510119.
- Dallas Card, Peter Henderson, Urvashi Khandelwal, Robin Jia, Kyle Mahowald, and Dan Jurafsky. 2020. **With little power comes great responsibility**. In *Proceedings of Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Dallas Card, Chenhao Tan, and Noah A. Smith. 2018. **Neural models for documents with metadata**. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2031–2040, Melbourne, Australia. Association for Computational Linguistics.
- Edward G Carmines, Michael J Ensley, and Michael W Wagner. 2012. Who fits the left-right divide? partisan polarization in the american electorate. *American Behavioral Scientist*, 56(12):1631–1653.
- Clifford J Carrubba, Matthew Gabel, Lacey Murrah, Ryan Clough, Elizabeth Montgomery, and Rebecca Schambach. 2006. Off the record: Unrecorded legislative votes, selection bias and roll-call vote analysis. *British Journal of Political Science*, 36(4):691–704.
- Thomas M Carsey and Geoffrey C Layman. 2006. Changing sides or changing minds? party identification and policy preferences in the american electorate. *American Journal of Political Science*, 50(2):464–477.
- Jamie L Carson, Gregory Koger, Matthew J Lebo, and Everett Young. 2010. The electoral costs of party loyalty in congress. *American Journal of Political Science*, 54(3):598–616.
- Adam Cayton. 2021. The policy substance of legislative ideology. *Legislative Studies Quarterly*, 46(3):679–712.
- Jonathan Chang, Jordan L. Boyd-Graber, Sean Gerrish, Chong Wang, and David M. Blei. 2009. **Reading tea leaves: How humans interpret topic models**. In *Proceedings of Advances in Neural Information Processing Systems*. Curran Associates, Inc.
- Xunru Che, Danaë Metaxa-Kakavouli, and Jeffrey T Hancock. 2018. Fake news in the news: An analysis of partisan coverage of the fake news phenomenon. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 289–292.
- Kaiping Chen, Amanda L Molder, Zening Duan, Shelley Boulianne, Christopher Eckart, Prince Mallari, and Diyi Yang. 2022. How climate movement actors

- and news media frame climate change and strike: Evidence from analyzing twitter and news media discourse from 2018 to 2021. *The International Journal of Press/Politics*, page 19401612221106405.
- Yimin Chen, Niall J Conroy, and Victoria L Rubin. 2015. Misleading online content: recognizing clickbait as "false news". In *Proceedings of the 2015 ACM on workshop on multimodal deception detection*, pages 15–19.
- Sophie Chesney, Maria Liakata, Massimo Poesio, and Matthew Purver. 2017. **Incongruent headlines: Yet another way to mislead your readers**. In *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism*, pages 56–61, Copenhagen, Denmark. Association for Computational Linguistics.
- Dennis Chong and James N Druckman. 2007. Framing theory. *Annu. Rev. Polit. Sci.*, 10:103–126.
- Jennifer Hayes Clark and Veronica Caro. 2013. Multimember districts and the substantive representation of women: An analysis of legislative cosponsorship networks. *Politics & Gender*, 9(1):1–30.
- Kevin Clark, Urvashi Khandelwal, Omer Levy, and Christopher D. Manning. 2019. **What does BERT look at? an analysis of BERT's attention**. In *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 276–286, Florence, Italy. Association for Computational Linguistics.
- Joshua Clinton, Simon Jackman, and Douglas Rivers. 2004. The statistical analysis of roll call data. *American Political Science Review*, 98(2):355–370.
- Joshua D Clinton, Anthony Bertelli, Christian R Grose, David E Lewis, and David C Nixon. 2012. Separated powers in the united states: The ideology of agencies, presidents, and congress. *American Journal of Political Science*, 56(2):341–354.
- Michael Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011. Political polarization on twitter. In *Proceedings of the international aaai conference on web and social media*, volume 5, pages 89–96.
- Alexander Coppock, Emily Ekins, David Kirby, et al. 2018. The long-lasting effects of newspaper op-eds on public opinion. *Quarterly Journal of Political Science*, 13(1):59–87.
- Lindsey Cormack. 2016. Extremity in congress: communications versus votes. *Legislative Studies Quarterly*, 41(3):575–603.
- Michael H Crespin and David W Rohde. 2010. Dimensions, issues, and bills: Appropriations voting on the house floor. *The Journal of Politics*, 72(4):976–989.

- James M Curry. 2015. *Legislating in the Dark: Information and Power in the House of Representatives*. University of Chicago Press.
- Robert A Dahl. 1956. *A preface to democratic theory*, volume 10. University of Chicago Press.
- Alina Darmstadt, Mick Prinz, and Oliver Saal. 2019. The murder of keira: Misinformation and hate speech as far-right online strategies.
- Maryam Davoodi, Eric Waltenburg, and Dan Goldwasser. 2020. Understanding the language of political agreement and disagreement in legislative texts. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5358–5368.
- Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. 2016. The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3):554–559.
- Dorottya Demszky, Nikhil Garg, Rob Voigt, James Zou, Jesse Shapiro, Matthew Gentzkow, and Dan Jurafsky. 2019. **Analyzing polarization in social media: Method and application to tweets on 21 mass shootings**. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2970–3005, Minneapolis, Minnesota. Association for Computational Linguistics.
- Matthew J Denny and Arthur Spirling. 2018. Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. In *Political Analysis*. Cambridge University Press.
- Bruce A Desmarais, Vincent G Moscardelli, Brian F Schaffner, and Michael S Kowal. 2015. Measuring legislative collaboration: The senate press events network. *Social Networks*, 40:43–54.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.
- Navid Dianati. 2016. Unwinding the hairball graph: Pruning algorithms for weighted complex networks. *Physical Review E*, 93(1):012304.
- Adji B Dieng, Francisco JR Ruiz, and David M Blei. 2020. Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, 8:439–453.

- Daniel Diermeier, Jean-François Godbout, Bei Yu, and Stefan Kaufmann. 2012. Language and ideology in congress. *British Journal of Political Science*, 42(1):31–55.
- Paul DiMaggio. 2015. Adapting computational text analysis to social science (and vice versa). *Big Data & Society*, 2(2):2053951715602908.
- Ran Ding, Ramesh Nallapati, and Bing Xiang. 2018. **Coherence-aware neural topic modeling**. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 830–836, Brussels, Belgium. Association for Computational Linguistics.
- Daniel DiSalvo. 2009. Party factions in congress. In *Congress & the presidency*, volume 36, pages 27–57. Taylor & Francis.
- Jesse Dodge, Suchin Gururangan, Dallas Card, Roy Schwartz, and Noah A. Smith. 2019. **Show your work: Improved reporting of experimental results**. In *Proceedings of Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Caitlin Doogan and Wray Buntine. 2021. **Topic model or topic twaddle? re-evaluating semantic interpretability measures**. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3824–3848, Online. Association for Computational Linguistics.
- Daniel Dor. 2003. On newspaper headlines as relevance optimizers. *Journal of pragmatics*, 35(5):695–721.
- William Downes and Senior Fellow William Downes. 1998. *Language and society*, volume 10. Cambridge university press.
- Lee Drutman. 2015. *The business of America is lobbying: How corporations became politicized and politics became more corporate*. Oxford University Press.
- JBrandon Duck-Mayr and Jacob Montgomery. 2022. Ends against the middle: Measuring latent traits when opposites respond the same way for antithetical reasons. *Political Analysis*, pages 1–20.
- Andrew Duffy, Edson Tandoc, and Rich Ling. 2020. Too good to be true, too good not to share: the social utility of fake news. *Information, Communication & Society*, 23(13):1965–1979.
- Eugenie Dugoua, Marion Dumas, and Joëlle Noailly. 2022. Text as data in environmental economics and policy. *Review of Environmental Economics and Policy*, 16(2):346–356.

- Daniel Ebanks, Hao Yan, R Michael Alvarez, Sanmay Das, and Betsy Sinclair. 2021. Leadership communication and power: Measuring leadership in the us house of representatives from social media data.
- Ullrich KH Ecker, Stephan Lewandowsky, Ee Pin Chang, and Rekha Pillai. 2014. The effects of subtle misinformation in news headlines. *Journal of experimental psychology: applied*, 20(4):323.
- Laura Edelson, Minh-Kha Nguyen, Ian Goldstein, Oana Goga, Damon McCoy, and Tobias Lauinger. 2021. Understanding engagement with us (mis) information news sources on facebook. pages 444–463.
- Patrick J Egan. 2013. *Partisan priorities: How issue ownership drives and distorts American politics*. Cambridge University Press.
- Jana Laura Egelhofer and Sophie Lecheler. 2019. Fake news as a two-dimensional phenomenon: A framework and research agenda. *Annals of the International Communication Association*, 43(2):97–116.
- Jacob Eisenstein, Amr Ahmed, and Eric P. Xing. 2011. **Sparse additive generative models of text**. In *Proceedings of the International Conference of Machine Learning*. Omnipress.
- Robert M Entman. 1993. Framing: Towards clarification of a fractured paradigm. *McQuail's reader in mass communication theory*, 390:397.
- Matthew Eshbaugh-Soha. 2005. The politics of presidential agendas. *Political Research Quarterly*, 58(2):257–268.
- Kevin M Esterling. 2007. Buying expertise: Campaign contributions and attention to policy analysis in congressional committees. *American Political Science Review*, 101(1):93–109.
- Kawin Ethayarajh and Dan Jurafsky. 2020. Utility is in the eye of the user: A critique of NLP leaderboard design. In *Proceedings of Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Norman Fairclough. 1993. Critical discourse analysis and the marketization of public discourse: The universities. *Discourse & society*, 4(2):133–168.
- Norman Fairclough. 2013. *Language and power*. Routledge.
- Justin Farrell. 2016. Corporate funding and ideological polarization about climate change. *Proceedings of the National Academy of Sciences*, 113(1):92–97.

- Stanley Feldman and Christopher Johnston. 2014. Understanding the determinants of political ideology: Implications of structural complexity. *Political Psychology*, 35(3):337–358.
- Jiachun Feng, Zusheng Zhang, Cheng Ding, Yanghui Rao, Haoran Xie, and Fu Lee Wang. 2022. Context reinforced neural topic modeling over short texts. *Information Sciences*, 607:79–91.
- Richard F Fenno. 2003. *Home style: House members in their districts*. Pearson College Division.
- Adriana Ferrugento, Hugo Gonalo Oliveira, Ana Alves, and Filipe Rodrigues. 2016. **Can topic modelling benefit from word sense information?** In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pages 3387–3393, Portoro, Slovenia. European Language Resources Association (ELRA).
- Mikhail Figurnov, S. Mohamed, and A. Mnih. 2018. Implicit reparameterization gradients. In *NeurIPS*.
- Manuel Fischer, Fr d ric Varone, Roy Gava, and Pascal Sciarini. 2019. **How MPs ties to interest groups matter for legislative co-sponsorship**. *Social Networks*, 57:34–42.
- Paul Frymer and Jacob M Grumbach. 2021. Labor unions and white racial politics. *American Journal of Political Science*, 65(1):225–240.
- Maksym Gabielkov, Arthi Ramachandran, Augustin Chaintreau, and Arnaud Legout. 2016. Social clicks: What and who gets read on twitter? pages 179–192.
- Saadia Gabriel, Skyler Hallinan, Maarten Sap, Pemi Nguyen, Franziska Roesner, Eunsol Choi, and Yejin Choi. 2022. Misinfo reaction frames: Reasoning about readers’ reactions to news headlines. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3108–3127.
- Ryan J Gallagher, Morgan R Frank, Lewis Mitchell, Aaron J Schwartz, Andrew J Reagan, Christopher M Danforth, and Peter Sheridan Dodds. 2021. Generalized word shift graphs: a method for visualizing and explaining pairwise comparisons between texts. *EPJ Data Science*, 10(1):4.
- Shuang Gao, Shivani Pandya, Smisha Agarwal, and Jo o Sedoc. 2021. **Topic modeling for maternal health using Reddit**. In *Proceedings of the 12th International Workshop on Health Text Mining and Information Analysis*, pages 69–76, online. Association for Computational Linguistics.

- John Gaventa. 1982. *Power and powerlessness: Quiescence and rebellion in an Appalachian valley*. University of Illinois Press.
- SoRelle Wyckoff Gaynor. 2021. The (financial) ties that bind: Social networks of intraparty caucuses. *Legislative Studies Quarterly*.
- Matthew Gentzkow, Bryan Kelly, and Matt Taddy. 2019. Text as data. *Journal of Economic Literature*, 57(3):535–74.
- Sean Gerrish and David Blei. 2012. How they vote: Issue-adjusted models of legislative behavior. *Advances in neural information processing systems*, 25:2753–2761.
- Sean M Gerrish and David M Blei. 2011. Predicting legislative roll calls from text. In *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*.
- Norjihan Abdul Ghani, Suraya Hamid, Ibrahim Abaker Targio Hashem, and Ejaz Ahmed. 2019. Social media big data analytics: A survey. *Computers in Human Behavior*, 101:417–428.
- Pranav Goel, Nikolay Malkin, SoRelle Gaynor, Nebojsa Jojic, Kristina Miler, and Philip Resnik. 2023. Donor activity is associated with us legislators’ attention to political issues. *Available at SSRN 4361565*.
- Matthew H Goldberg, Jennifer R Marlon, Xinran Wang, Sander van der Linden, and Anthony Leiserowitz. 2020. Oil and gas companies invest in legislators that vote against the environment. *Proceedings of the National Academy of Sciences*, 117(10):5111–5112.
- J David Gopoian, Hobart Smith, and William Smith. 1984. What makes pacs tick? an analysis of the allocation patterns of economic interest groups. *American Journal of Political Science*, pages 259–281.
- Jianping Gou, Baosheng Yu, Stephen J Maybank, and Dacheng Tao. 2021. Knowledge distillation: A survey. *International Journal of Computer Vision*, 129:1789–1819.
- Donald Philip Green and Jonathan S Krasno. 1988. Salvation for the spendthrift incumbent: Reestimating the effects of campaign spending in house elections. *American Journal of Political Science*, pages 884–907.
- Jon Green, Stefan McCabe, Sarah Shugars, John Harrington, Hanyu Chwe, Shuyang Cao, and David Lazer. 2021. Curation bubbles: Domain versus url level analysis of partisan news sharing on social media. *Working Paper*. <https://osf.io/udfaz/>.
- Michael Gregory and Susanne Carroll. 2018. *Language and situation: Language varieties and their social contexts*. Routledge.

- Kevin B Grier and Michael C Munger. 1991. Committee assignments, constituent preferences, and campaign contributions. *Economic Inquiry*, 29(1):24–43.
- Thomas L Griffiths and Mark Steyvers. 2004. Finding scientific topics. In *Proceedings of the National Academy of Sciences*. National Academy of Sciences.
- Tom Griffiths. 2002. Gibbs sampling in the generative model of latent dirichlet allocation. 518.
- Justin Grimmer. 2013. *Representational style in Congress: What legislators say and why it matters*. Cambridge University Press.
- Justin Grimmer, Margaret E Roberts, and Brandon M Stewart. 2022. *Text as data: A new framework for machine learning and the social sciences*. Princeton University Press.
- Justin Grimmer and Brandon M Stewart. 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 21(3):267–297.
- Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. Fake news on twitter during the 2016 us presidential election. *Science*, 363(6425):374–378.
- Justin H Gross and Justin H Kirkland. 2019. Rivals or allies? a multilevel analysis of cosponsorship within state delegations in the us senate. In *Congress & the Presidency*, volume 46, pages 183–213. Taylor & Francis.
- Yupeng Gu, Ting Chen, Yizhou Sun, and Bingyu Wang. 2016. Ideology detection for twitter users with heterogeneous types of links. *arXiv preprint arXiv:1612.08207*.
- Yupeng Gu, Yizhou Sun, Ning Jiang, Bingyu Wang, and Ting Chen. 2014. Topic-factorized ideal point estimation model for legislative voting network. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 183–192. ACM.
- Andrew Guess, Jonathan Nagler, and Joshua Tucker. 2019. Less than you think: Prevalence and predictors of fake news dissemination on facebook. *Science advances*, 5(1):eaau4586.
- Andrew M Guess. 2021. (almost) everything in moderation: New evidence on americans’ online media diets. *American Journal of Political Science*, 65(4):1007–1022.

- Andy Guess, Kevin Aslett, Joshua Tucker, Richard Bonneau, and Jonathan Nagler. 2021. Cracking open the news feed: Exploring what us facebook users see and share with large-scale platform data. *Journal of Quantitative Description: Digital Media*, 1.
- Lin Gui, Jia Leng, Gabriele Pergola, Yu Zhou, Ruifeng Xu, and Yulan He. 2019. **Neural topic model with reinforcement learning**. In *Proceedings of Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Lei Guo and Chris Vargo. 2020. “fake news” and emerging online media ecosystem: An integrated intermedia agenda-setting analysis of the 2016 us presidential election. *Communication Research*, 47(2):178–200.
- Deepak Gupta and Rinkle Rani. 2019. A study of big data evolution and research challenges. *Journal of information science*, 45(3):322–340.
- Pankaj Gupta, Yatin Chaudhary, F. Buettner, and Hinrich Schütze. 2019a. **textTOvec: Deep contextualized neural autoregressive models of language with distributed compositional prior**. In *Proceedings of the International Conference on Learning Representations*.
- Pankaj Gupta, Yatin Chaudhary, Florian Buettner, and Hinrich Schütze. 2019b. **Document informed neural autoregressive topic models with distributional prior**. In *Association for the Advancement of Artificial Intelligence*. AAAI Press.
- Louisa Ha, Loarre Andreu Perez, and Rik Ray. 2021. Mapping recent development in scholarship on fake news and misinformation, 2008 to 2017: Disciplinary contribution, topics, and impact. *American behavioral scientist*, 65(2):290–315.
- Jacob S Hacker and Paul Pierson. 2005. Abandoning the middle: The bush tax cuts and the limits of democratic control. *Perspectives on Politics*, 3(1):33–53.
- Andrew B Hall. 2015. What happens when extremists win primaries? *American Political Science Review*, 109(1):18–42.
- Richard L Hall. 1987. Participation and purpose in committee decision making. *American Political Science Review*, 81(1):105–127.
- Richard L Hall and Alan V Deardorff. 2006. Lobbying as legislative subsidy. *American Political Science Review*, 100(1):69–84.
- Richard L Hall and Frank W Wayman. 1990. Buying time: Moneyed interests and the mobilization of bias in congressional committees. *American political science review*, 84(3):797–820.

- Michael Hameleers, Toni van der Meer, and Rens Vliegenthart. 2022. Civilized truths, hateful lies? incivility and hate speech in false information—evidence from fact-checked statements in the us. *Information, Communication & Society*, 25(11):1596–1613.
- Paul J Hanges, Frances Lee, Kristina Miler, and Jennifer Wessel. 2020. Report on the organizational climates of congress. *University of Maryland*.
- John Mark Hansen. 1991. *Gaining access: Congress and the farm lobby, 1919-1981*. University of Chicago Press.
- Wendy L Hansen, Michael S Rocca, and Brittany Leigh Ortiz. 2015. The effects of citizens united on corporate spending in the 2012 presidential election. *The Journal of Politics*, 77(2):535–545.
- Douglas B Harris. 2005. Orchestrating party talk: A party-based view of one-minute speeches in the house of representatives. *Legislative Studies Quarterly*, 30(1):127–141.
- Joseph C Harry. 2014. Journalistic quotation: Reported speech in newspapers from a semiotic-linguistic perspective. *Journalism*, 15(8):1041–1058.
- Todd K Hartman, Thomas VA Stocks, Ryan McKay, Jilly Gibson-Miller, Liat Levita, Anton P Martinez, Liam Mason, Orla McBride, Jamie Murphy, Mark Shevlin, et al. 2021. The authoritarian dynamic during the covid-19 pandemic: Effects on nationalism and anti-immigrant sentiment. *Social Psychological and Personality Science*, 12(7):1274–1285.
- Hans JG Hassell. 2016. Party control of party primaries: Party influence in nominations for the us senate. *The Journal of Politics*, 78(1):75–87.
- Carina I Hausladen, Marcel H Schubert, and Elliott Ash. 2020. Text classification of ideological direction in judicial opinions. *International Review of Law and Economics*, 62:105903.
- Ruifang He, Xuefei Zhang, Di Jin, Longbiao Wang, Jianwu Dang, and Xiangang Li. 2018. **Interaction-aware topic model for microblog conversations through network embedding and user attention**. In *International Conference on Computational Linguistics*. Association for Computational Linguistics.
- Libby Hemphill and Matthew A Shapiro. 2019. Appealing to the base or to the moveable middle? incumbents’ partisan messaging before the 2016 us congressional elections. *Journal of Information Technology & Politics*, 16(4):325–341.
- Alexander Hertel-Fernandez. 2014. Who passes business’s “model bills”? policy capacity and corporate influence in us state politics. *Perspectives on Politics*, 12(3):582–602.

- Alexander Hertel-Fernandez. 2018. *Politics at Work: How Companies Turn Their Workers into Lobbyists*. Oxford University Press.
- Alexander Hertel-Fernandez. 2019. *State capture: how conservative activists, big businesses, and wealthy donors reshaped the American states—and the nation*. Oxford University Press, USA.
- Benjamin Highton. 2019. issue accountability in us house elections. *Political Behavior*, 41(2):349–367.
- Benjamin Highton and Michael S Rocca. 2005. Beyond the roll-call arena: The determinants of position taking in congress. *Political Research Quarterly*, 58(2):303–316.
- Kim Quaile Hill. 2001. Multiple-method measurement of legislators’ ideologies. *Legislative Studies Quarterly*, pages 263–274.
- Geoffrey Hinton, Oriol Vinyals, and Jeffrey Dean. 2015. Distilling the knowledge in a neural network. In *NIPS Deep Learning and Representation Learning Workshop*.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. **Long Short-Term Memory**. *Neural Computation*, 9(8):1735–1780.
- Thomas Hofmann. 2001. Unsupervised learning by probabilistic latent semantic analysis. *Machine learning*, 42(1):177–196.
- Marie Hojnacki and David C Kimball. 1998. Organized interests and the decision of whom to lobby in congress. *American Political Science Review*, 92(4):775–790.
- Paul W Holland and Samuel Leinhardt. 1981. An exponential family of probability distributions for directed graphs. *Journal of the american Statistical association*, 76(373):33–50.
- Liangjie Hong, Aziz S Doumith, and Brian D Davison. 2013. Co-factorization machines: modeling user interests and predicting individual decisions in twitter. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 557–566. ACM.
- Sounman Hong and Sun Hyoungh Kim. 2016. Political polarization on twitter: Implications for the use of social media in digital governments. *Government Information Quarterly*, 33(4):777–782.
- Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. **spaCy: Industrial-strength Natural Language Processing in Python**.

- Alexander Hoyle, Pranav Goel, Andrew Hian-Cheong, Denis Peskov, Jordan Boyd-Graber, and Philip Resnik. 2021. **Is automated topic model evaluation broken? the incoherence of coherence**. In *Advances in Neural Information Processing Systems*, volume 34, pages 2018–2033. Curran Associates, Inc.
- Alexander Hoyle, Rupak Sarkar, Pranav Goel, and Philip Resnik. 2023. Making the implicit explicit: Implicit content as a first class citizen in nlp. *arXiv preprint arXiv:2305.14583*.
- Alexander Miserlis Hoyle, Pranav Goel, and Philip Resnik. 2020. **Improving Neural Topic Models using Knowledge Distillation**. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1752–1771, Online. Association for Computational Linguistics.
- Alexander Miserlis Hoyle, Rupak Sarkar, Pranav Goel, and Philip Resnik. 2022. **Are neural topic models broken?** In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 5321–5344, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- Hengtong Hu, Lingxi Xie, Richang Hong, and Qi Tian. 2020a. Creating Something from Nothing: Unsupervised Knowledge Distillation for Cross-Modal Hashing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3123–3132.
- Weihua Hu and Jun’ichi Tsujii. 2016. A latent concept topic model for robust topic inference using word embeddings. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 380–386.
- Xuemeng Hu, Rui Wang, Deyu Zhou, and Yuxuan Xiong. 2020b. **Neural topic modeling with cycle-consistent adversarial training**. In *Proceedings of Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Mingkun Huang, Yongbin You, Zhehuai Chen, Yanmin Qian, and Kai Yu. 2018. **Knowledge Distillation for Sequence Model**. In *Interspeech 2018*, pages 3703–3707. ISCA.
- Adam G Hughes, Stefan D McCabe, William R Hobbs, Emma Remy, Sono Shah, and David MJ Lazer. 2021. Using administrative records and survey data to construct samples of tweeters and tweets. *Public Opinion Quarterly*, 85(S1):323–346.
- Edda Humprecht. 2019. Where ‘fake news’ flourishes: a comparison across four western democracies. *Information, Communication & Society*, 22(13):1973–1988.
- Iliyan R Iliev. 2021. The power dynamics of campaign contributions and legislative rhetoric. *Interest Groups & Advocacy*, 10(3):240–263.

- Masaru Isonuma, Junichiro Mori, Danushka Bollegala, and Ichiro Sakata. 2020. **Tree-Structured Neural Topic Model**. In *Proceedings of the Association for Computational Linguistics*. Association for Computational Linguistics.
- Shanto Iyengar and Adam Simon. 1993. News coverage of the gulf crisis and public opinion: A study of agenda-setting, priming, and framing. *Communication research*, 20(3):365–383.
- John E Jackson and John W Kingdon. 1992. Ideology, interest group scores, and legislative votes. *American Journal of Political Science*, pages 805–823.
- Gary C Jacobson and Jamie L Carson. 2019. *The politics of congressional elections*. Rowman & Littlefield.
- Martin Jankowiak and Fritz Obermeyer. 2018a. **Pathwise derivatives beyond the reparameterization trick**. In *Proceedings of the International Conference of Machine Learning*. PMLR.
- Martin Jankowiak and Fritz Obermeyer. 2018b. Pathwise Derivatives Beyond the Reparameterization Trick. In *International Conference on Machine Learning*, pages 2235–2244.
- Bryan D Jones and Frank R Baumgartner. 2005. *The politics of attention: How government prioritizes problems*. University of Chicago Press.
- Weonyoung Joo, Wonsung Lee, Sungrae Park, and Il-Chul Moon. 2020. **Dirichlet variational autoencoder**. *Pattern Recognition*, 107:107514.
- John T Jost and Joanna Sterling. 2020. The language of politics: ideological differences in congressional communication on social media and the floor of congress. *Social Influence*, 15(2-4):80–103.
- Nicholas Judd, Dan Drinkard, Jeremy Carbaugh, and Lindsay Young. 2017. congressional-record: A parser for the Congressional Record. Chicago, IL.
- Namkyu Jung and Hyeong In Choi. 2017. Continuous semantic topic embedding model using variational autoencoder. *arXiv e-prints*, pages arXiv–1711.
- Joshua L Kalla and David E Broockman. 2016. Campaign contributions facilitate access to congressional officials: A randomized field experiment. *American Journal of Political Science*, 60(3):545–558.
- Katharina Kann, Sascha Rothe, and Katja Filippova. 2018. **Sentence-level fluency evaluation: References help, but can be spared!** In *Proceedings of the 22nd Conference on Computational Natural Language Learning*, pages 313–323, Brussels, Belgium. Association for Computational Linguistics.

- Lisa Kelly. 2010. Public personas, private lives and the power of the celebrity comedian: a consideration of the ross and brand ‘sachsgate’ affair. *Celebrity Studies*, 1(1):115–117.
- In Song Kim, John Londregan, and Marc Ratkovic. 2018. Estimating spatial preferences from votes and text. *Political Analysis*, 26(2):210–229.
- Jae Yeon Kim and Aniket Kesari. 2021. Misinformation and hate speech: The case of anti-asian hate speech during the covid-19 pandemic. *Journal of Online Trust and Safety*, 1(1).
- Yoon Kim and Alexander M. Rush. 2016. **Sequence-level knowledge distillation**. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1317–1327, Austin, Texas. Association for Computational Linguistics.
- John W Kingdon. 1989. *Congressmen’s voting decisions*. University of Michigan Press.
- Diederik P. Kingma and Jimmy Ba. 2015. **Adam: A method for stochastic optimization**. In *Proceedings of the International Conference on Learning Representations*.
- Diederik P. Kingma and Max Welling. 2014. **Auto-Encoding Variational Bayes**. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*.
- Justin H Kirkland. 2011. The relational determinants of legislative outcomes: Strong and weak ties between legislators. *The Journal of Politics*, 73(3):887–898.
- Justin H Kirkland. 2012. Multimember districts’ effect on collaboration between us state legislators. *Legislative Studies Quarterly*, 37(3):329–353.
- Anastassia Kornilova, Daniel Argyle, and Vladimir Eidelman. 2018. Party matters: Enhancing legislative embeddings with author attributes for vote prediction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 510–515.
- Elitsa Kortenska, Bernard Steunenberg, and Indraneel Sircar. 2020. Public-elite gap on european integration: the missing link between discourses among citizens and elites in serbia. *Journal of European integration*, 42(6):873–888.
- Neema Kotonya and Francesca Toni. 2020. Explainable automated fact-checking for public health claims. pages 7740–7754.
- Peter Kraft, Hirsh Jain, and Alexander M. Rush. 2016. **An Embedding Model for Predicting Roll-Call Votes**. pages 2066–2070.

- Katsiaryna Krasnashchok and Salim Jouili. 2018. **Improving topic quality by promoting named entities in topic modeling**. In *Proceedings of the Association for Computational Linguistics*. Association for Computational Linguistics.
- Keith Krehbiel. 1993. Where's the party? *British Journal of Political Science*, 23(2):235–266.
- Randall S Kroszner and Thomas Stratmann. 2005. Corporate campaign contributions, repeat giving, and the rewards to legislator reputation. *The Journal of Law and Economics*, 48(1):41–71.
- Emily Kubin and Christian von Sikorski. 2021. The role of (social) media in political polarization: a systematic review. *Annals of the International Communication Association*, 45(3):188–206.
- Ken Lang. 1995. **NewsWeeder: Learning to Filter Netnews**. In Armand Prieditis and Stuart Russell, editors, *Machine Learning Proceedings 1995*, pages 331–339. Morgan Kaufmann, San Francisco (CA).
- Laura I Langbein. 1986. Money and access: Some empirical evidence. *The journal of politics*, 48(4):1052–1062.
- Timothy M LaPira and Herschel F Thomas. 2017. *Revolving door lobbying: Public service, private influence, and the unequal representation of interests*. University Press of Kansas.
- Jey Han Lau, David Newman, and Timothy Baldwin. 2014. **Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality**. In *Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics.
- Benjamin E Lauderdale and Tom S Clark. 2014. Scaling politically meaningful dimensions using texts and votes. *American Journal of Political Science*, 58(3):754–771.
- Benjamin E Lauderdale and Alexander Herzog. 2016. Measuring political positions from legislative speech. *Political Analysis*, 24(3):374–394.
- Michael Laver, Kenneth Benoit, and John Garry. 2003. Extracting policy positions from political texts using words as data. *American political science review*, 97(2):311–331.
- David Lazer and Jason Radford. 2017. Data ex machina: introduction to big data. *Annual Review of Sociology*, 43:19–39.

- David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. 2018. The science of fake news. *Science*, 359(6380):1094–1096.
- Tak Yeon Lee, Alison Smith, Kevin Seppi, Niklas Elmqvist, Jordan Boyd-Graber, and Leah Findlater. 2017. The human touch: How non-expert users perceive, interpret, and fix topic models. *International Journal of Human-Computer Studies*, 105:28–42.
- Joshua Y Lerner and Robert Shaffer. 2020. Setting the committee agenda: Measuring speaker influence in congressional hearings.
- Hause Lin, Jana Lasser, Stephan Lewandowsky, Rocky Cole, Andrew Gully, David G Rand, and Gordon Pennycook. 2022. **High level of agreement across different news domain quality ratings**.
- Jianhua Lin. 1991. Divergence measures based on the shannon entropy. *IEEE Transactions on Information theory*, 37(1):145–151.
- Lihui Lin, Hongyu Jiang, and Yanghui Rao. 2020. **Copula guided neural topic modelling for short texts**. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM.
- Tianyi Lin, Zhiyue Hu, and Xin Guo. 2019. **Sparsemax and relaxed wasserstein for topic sparsity**. In *International Conference on Web Search and Data Mining (WSDM)*. ACM.
- Zachary C Lipton. 2018. **The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery**. In *Queue*. ACM.
- Luyang Liu, Heyan Huang, Yang Gao, Yongfeng Zhang, and Xiaochi Wei. 2019a. **Neural variational correlated topic modeling**. In *Proceedings of the World Wide Web Conference*. ACM.
- Nelson F. Liu, Matt Gardner, Yonatan Belinkov, Matthew E. Peters, and Noah A. Smith. 2019b. **Linguistic knowledge and transferability of contextual representations**. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1073–1094, Minneapolis, Minnesota. Association for Computational Linguistics.
- Xuan Liu, Xiaoguang Wang, and Stan Matwin. 2018a. Improving the Interpretability of Deep Neural Networks with Knowledge Distillation. In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 905–912. IEEE.

- Yongcheng Liu, Lu Sheng, Jing Shao, Junjie Yan, Shiming Xiang, and Chunhong Pan. 2018b. **Multi-Label Image Classification via Knowledge Distillation from Weakly-Supervised Detection**. *2018 ACM Multimedia Conference on Multimedia Conference - MM '18*, pages 700–708.
- Alessandro Lovari. 2020. Spreading (dis) trust: Covid-19 misinformation and government intervention in Italy. *Media and Communication*, 8(2):458–461.
- Will Lowe, Kenneth Benoit, Slava Mikhaylov, and Michael Laver. 2011. Scaling policy preferences from coded political texts. *Legislative studies quarterly*, 36(1):123–155.
- Jeffrey Lund, Piper Armstrong, Wilson Fearn, Stephen Cowley, Emily Hales, and Kevin Seppi. 2019. **Cross-referencing using fine-grained topic modeling**. In *Proceedings of the Association for Computational Linguistics*, Minneapolis, Minnesota. Association for Computational Linguistics.
- Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 2011. **Learning word vectors for sentiment analysis**. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 142–150, Portland, Oregon, USA. Association for Computational Linguistics.
- Seraphine F Maerz and Cornelius Puschmann. 2020. Text as data for conflict research: A literature survey. *Computational Conflict Research*, pages 43–65.
- Sílvia Majó-Vázquez, Rasmus K Nielsen, and Sandra González-Bailón. 2019. The backbone structure of audience networks: A new approach to comparing online news consumption across countries. *Political Communication*, 36(2):227–240.
- Forrest Maltzman and Lee Sigelman. 1996. The politics of talk: Unconstrained floor time in the US House of Representatives. *The Journal of Politics*, 58(3):819–830.
- Henry Berthold Mann and Donald Ransom Whitney. 1947. **On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other**. In *The Annals of Mathematical Statistics*. Institute of Mathematical Statistics.
- Stephen Marche. 2012. **Literature is not data: Against digital humanities**. In *LA Review of Books*.
- Donald W Marquardt. 1970. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, 12(3):591–612.
- Alice E Marwick and Rebecca Lewis. 2017. Media manipulation and disinformation online.

- Nitika Mathur, Timothy Baldwin, and Trevor Cohn. 2020. **Tangled up in BLEU: Reevaluating the evaluation of automatic machine translation evaluation metrics**. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4984–4997, Online. Association for Computational Linguistics.
- Donald R Matthews. 1959. The folkways of the united states senate: Conformity to group norms and legislative effectiveness. *American Political Science Review*, 53(4):1064–1089.
- Donald R Matthews and James A Stimson. 1975. *Yeas and nays: Normal decision-making in the US House of Representatives*. Wiley-Interscience.
- David Mayhew. 1974. *Congress: The electoral connection*. new haven, ct: Yale university press.
- David R Mayhew. 2004. *Congress: The electoral connection*. Yale university press.
- Marc J. Mazerolle. 2023. **Aiccmodavg: Model selection and multimodel inference based on (q)aic(c)**. R package version 2.3.2.
- Andrew Kachites McCallum. 2002. Machine learning with MALLET: A Machine Learning for Language Toolkit. <http://mallet.cs.umass.edu>.
- Maxwell McCombs. 1994. The future agenda for agenda setting research. *Journal of mass communication studies*, 45:171–181.
- Maxwell McCombs. 2002. The agenda-setting role of the mass media in the shaping of public opinion. In *Mass Media Economics 2002 Conference, London School of Economics*: <http://sticerd.lse.ac.uk/dps/extra/McCombs.pdf>.
- Maxwell McCombs and Salma I Ghanem. 2001. The convergence of agenda setting and framing. In *Framing public life*, pages 83–98. Routledge.
- Matthew T McCrudden and Gregory Schraw. 2007. Relevance and goal-focusing in text processing. *Educational psychology review*, 19:113–139.
- Amy Melissa McKay. 2018. Fundraising for favors? linking lobbyist-hosted fundraisers to legislative benefits. *Political Research Quarterly*, 71(4):869–880.
- Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444.
- Florian Meier and David Elswailer. 2019. **Studying politicians’ information sharing on social media**. *CHIIR 2019 - Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, (March):237–241.

- I Dan Melamed. 2000. Models of translational equivalence among words. *Computational Linguistics*, 26(2):221–249.
- Stephen Merity, Caiming Xiong, James Bradbury, and Richard Socher. 2017. **Pointer Sentinel Mixture Models**. *ICLR*.
- P. T. Metaxas, E. Mustafaraj, K. Wong, L. Zeng, M. O’Keefe, and S. Finn. 2015. What do retweets indicate? Results from user survey and meta-review of research. *Proceedings of the 9th International Conference on Web and Social Media, ICWSM 2015*, pages 658–661.
- Yishu Miao, Edward Grefenstette, and Phil Blunsom. 2017. **Discovering discrete latent topics with neural variational inference**. In *Proceedings of the International Conference of Machine Learning*. PMLR.
- Yishu Miao, Lei Yu, and Phil Blunsom. 2016. **Neural Variational Inference for Text Processing**. *ICML*.
- Kristina Miler. 2017. How committees shape legislative behavior: An examination of interests and institutions. *American Politics Research*, 45(5):813–839.
- Kristina C Miler. 2018. *Poor representation: Congress and the politics of poverty in the United States*. Cambridge University Press.
- Matt Miller. 2016. Climate change acceptors in the republican party: A model of the determinants of party defection.
- Warren E Miller and Donald E Stokes. 1963. Constituency influence in congress. *American political science review*, 57(1):45–56.
- David Mimno, Hanna Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. **Optimizing semantic coherence in topic models**. In *Proceedings of Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Katarzyna Molek-Kozakowska. 2013. Towards a pragma-linguistic framework for the study of sensationalism in news headlines. *Discourse & Communication*, 7(2):173–197.
- Burt L Monroe, Michael P Colaresi, and Kevin M Quinn. 2008. Fightin’ words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*, 16(4):372–403.
- Corrado Monti, Giuseppe Manco, Cigdem Aslay, and Francesco Bonchi. 2021. Learning ideological embeddings from information cascades. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 1325–1334.

- Fred Morstatter and Huan Liu. 2018. **In search of coherence and consensus: Measuring the interpretability of statistical topics**. *Journal of Machine Learning Research*.
- Scott Moser, Abel Rodríguez, and Chelsea L Lofland. 2021. Multiple ideal points: Revealed preferences in different domains. *Political Analysis*, 29(2):139–166.
- Finn Müller-Hansen, Max W Callaghan, and Jan C Minx. 2020. Text as big data: Develop codes of practice for rigorous computational text analysis in energy social science. *Energy Research & Social Science*, 70:101691.
- Feng Nan, Ran Ding, Ramesh Nallapati, and Bing Xiang. 2019. **Topic modeling with Wasserstein autoencoders**. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 6345–6381, Florence, Italy. Association for Computational Linguistics.
- Thomas E Nelson, Rosalee A Clawson, and Zoe M Oxley. 1997. Media framing of a civil liberties conflict and its effect on tolerance. *American Political Science Review*, 91(3):567–583.
- David Newman, Jey Han Lau, Karl Grieser, and Timothy Baldwin. 2010. **Automatic evaluation of topic coherence**. In *Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics.
- NewsGuard. 2021. Newsguard. <https://www.newsguardtech.com/>.
- Dat Quoc Nguyen, Richard Billingsley, Lan Du, and Mark Johnson. 2015a. **Improving topic models with latent feature word representations**. *Transactions of the Association for Computational Linguistics*.
- Viet-An Nguyen, Jordan Boyd-Graber, Philip Resnik, and Kristina Miler. 2015b. **Tea party in the house: A hierarchical ideal point topic model and its application to republican legislators in the 112th congress**. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1438–1448, Beijing, China. Association for Computational Linguistics.
- Viet-An Nguyen, Jordan L Ying, and Philip Resnik. 2013. Lexical and hierarchical topic regression. In *Advances in neural information processing systems*, pages 1106–1114.
- Xuefei Ning, Y. Zheng, Zhuxi Jiang, Y. Wang, H. Yang, and J. Huang. 2020. Non-parametric topic modeling with neural inference. In *Neurocomputing*.

- Katherine Ognyanova, David Lazer, Ronald E Robertson, and Christo Wilson. 2020. Misinformation in action: Fake news exposure is linked to lower trust in media, higher trust in government when your side is in power. *Harvard Kennedy School Misinformation Review*.
- Kieron O'Hara and David Stevens. 2015. Echo chambers and online radicalism: Assessing the internet's complicity in violent extremism. *Policy & Internet*, 7(4):401–422.
- Tracy Osborn, Rebecca J Kreitzer, Emily U Schilling, and Jennifer Hayes Clark. 2019. Ideology and polarization among women state legislators. *Legislative Studies Quarterly*, 44(4):647–680.
- Tracy Osborn and Jeanette Morehouse Mendez. 2010. Speaking as women: Women and floor speeches in the senate. *Journal of Women, Politics & Policy*, 31(1):1–21.
- José Otero and Walter Kintsch. 1992. Failures to detect contradictions in a text: What readers believe versus what they read. *Psychological Science*, 3(4):229–236.
- Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744.
- Madhur Panwar, Shashank Shailabh, Milan Aggarwal, and Balaji Krishnamurthy. 2020. TAN-NTM: Topic attention networks for neural topic modeling. In *Proceedings of the Association for Computational Linguistics*.
- David CW Parker and Craig Goodman. 2009. Making a good impression: Resource allocation, home styles, and washington work. *Legislative Studies Quarterly*, 34(4):493–524.
- Robert Parker, David Graff, Junbo Kong, Ke Chen, and Kazuaki Maeda. 2011. English Gigaword Fifth Edition.
- Scott D. Pauls, Greg Leibon, and Daniel Rockmore. 2015. **The social identity voting model: Ideology and community structures**. *Research and Politics*, 2(2).
- Kathryn Pearson and Logan Dancey. 2011a. Elevating women's voices in congress: Speech participation in the house of representatives. *Political Research Quarterly*, 64(4):910–923.
- Kathryn Pearson and Logan Dancey. 2011b. Speaking for the underrepresented in the house of representatives: Voicing women's interests in a partisan era. *Politics & Gender*, 7(4):493–519.

- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Min Peng, Qianqian Xie, Yanchun Zhang, Hua Wang, Xiuzhen Zhang, Jimin Huang, and Gang Tian. 2018. **Neural sparse topical coding**. In *Proceedings of the Association for Computational Linguistics*. Association for Computational Linguistics.
- Tai-Quan Peng, Mengchen Liu, Yingcai Wu, and Shixia Liu. 2016. Follower-follower network, communication networks, and vote agreement of the us members of congress. *Communication Research*, 43(7):996–1024.
- Nathaniel Persily. 2017. The 2016 us election: Can democracy survive the internet? *Journal of democracy*, 28(2):63–76.
- John R Petrocik. 1996. Issue ownership in presidential elections, with a 1980 case study. *American journal of political science*, pages 825–850.
- Paul Pierson. 2017. American hybrid: Donald trump and the strange merger of populism and plutocracy. *The British journal of sociology*, 68:S105–S119.
- Andrew Piper, Richard Jean So, and David Bamman. 2021. Narrative theory for computational narrative understanding. pages 298–311.
- Keith T Poole and Howard Rosenthal. 1985. A spatial model for legislative roll call analysis. *American journal of political science*, pages 357–384.
- Keith T Poole and Howard Rosenthal. 1991. Patterns of congressional voting. *American journal of political science*, pages 228–278.
- Keith T Poole and Howard Rosenthal. 2000. *Congress: A political-economic history of roll call voting*. Oxford University Press on Demand.
- Bruno Pouliquen, Ralf Steinberger, and Clive Best. 2007. Automatic detection of quotations in multilingual news. In *Proceedings of Recent Advances in Natural Language Processing*, pages 487–492.
- Eleanor Neff Powell and Justin Grimmer. 2016. Money in exile: Campaign contributions and committee access. *The Journal of Politics*, 78(4):974–988.
- Lynda W Powell. 2013. The influence of campaign contributions on legislative policy. In *The Forum*, volume 11, pages 339–355. De Gruyter.
- John W Pratt. 1959. Remarks on zeros and ties in the wilcoxon signed rank procedures. *Journal of the American Statistical Association*, 54(287):655–667.

- Stephen Quinlan, Tobias Gummer, Joss Roßmann, and Christof Wolf. 2018. ‘show me the money and the party!’—variation in facebook and twitter adoption by politicians. *Information, communication & society*, 21(8):1031–1049.
- Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(140):1–67.
- Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the Language Resources and Evaluation Conference*. ELRA.
- Beth Reingold. 2008. Women as office holders: Linking descriptive and substantive representation. *Political women and American democracy*, 9:128–147.
- Lisa Reppell and Erica Shein. 2019. Disinformation campaigns and hate speech: Exploring the relationship and programming interventions. *Arlington, VA: International Foundation for Electoral Systems*.
- Mehdi Rezaee and Francis Ferraro. 2020. **A discrete variational recurrent topic model without the reparametrization trick**. In *Proceedings of Advances in Neural Information Processing Systems*. Curran Associates, Inc.
- Nils Ringe, Jennifer Nicoll Victor, and Justin H Gross. 2013. Keeping your friends close and your enemies closer? information networks in legislative politics. *British Journal of Political Science*, 43(3):601–628.
- Margaret E Roberts, Brandon M Stewart, Dustin Tingley, Edoardo M Airolidi, et al. 2013. The structural topic model and applied social science. In *Advances in neural information processing systems workshop on topic models: computation, application, and evaluation*, volume 4, pages 1–20. Harrahs and Harveys, Lake Tahoe.
- Margaret E Roberts, 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.
- Craig T Robertson and Rachel R Mourão. 2020. Faking alternative journalism? an analysis of self-presentations of “fake news” sites. *Digital Journalism*, 8(8):1011–1029.
- Michael S Rocca. 2007. Nonlegislative debate in the us house of representatives. *American Politics Research*, 35(4):489–505.
- Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. **Exploring the space of topic coherence measures**. In *International Conference on Web Search and Data Mining (WSDM)*. ACM.

- Joshua Roesslein. 2020. Tweepy: Twitter for python. URL: <https://github.com/tweepy/tweepy>, 484.
- Anna Rogers, Olga Kovaleva, and Anna Rumshisky. 2020. A primer in bertology: What we know about how bert works. *Transactions of the Association for Computational Linguistics*, 8:842–866.
- Kaleigh Rogers. 2019. **The latest disinformation threat online? old news stories.** *CBC News*.
- Thomas Romer and James M Snyder Jr. 1994. An empirical investigation of the dynamics of pac contributions. *American journal of political science*, pages 745–769.
- Jon Roozenbeek and Sander Van Der Linden. 2019. The fake news game: actively inoculating against the risk of misinformation. *Journal of risk research*, 22(5):570–580.
- Douglas D Roscoe and Shannon Jenkins. 2005. A meta-analysis of campaign contributions’ impact on roll call voting. *Social Science Quarterly*, 86(1):52–68.
- Annelise Russell. 2018. Us senators on twitter: Asymmetric party rhetoric in 140 characters. *American Politics Research*, 46(4):695–723.
- Annelise Russell. 2021. Gendered priorities? policy communication in the us senate. In *Congress & the Presidency*, volume 48, pages 319–342. Taylor & Francis.
- A. Samuels. 2022. **Why republicans in blue cities are increasingly outliers.**
- Evan Sandhaus. 2008. The New York Times annotated corpus. In *Linguistic Data Consortium*.
- Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. **DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter.** *NeurIPS EMC² Workshop*.
- Elmer Eric Schattschneider. 1960. *The semisovereign people: a realist’s view of democracy in America*. Holt, Rinehart and Winston.
- Kay Lehman Schlozman, Sidney Verba, and Henry E Brady. 2012. *The unheavenly chorus: Unequal political voice and the broken promise of American democracy*. Princeton University Press.
- Benjamin M Schmidt. 2012. **Words alone: Dismantling topic models in the humanities.** In *Journal of Digital Humanities*.
- Alexandra Schofield and David Mimno. 2016. **Comparing apples to apple: The effects of stemmers on topic models.** *Transactions of the Association for Computational Linguistics*.

- Matthew A Shapiro, Libby Hemphill, and Jahna Otterbacher. 2012. Doing what i say: connecting congressional social media behavior and congressional voting. In *Proceedings of the Annual Meeting of the Midwest Political Science Association*, pages 1–32. Midwest Political Science Association Chicago.
- Or Sharir, Barak Peleg, and Yoav Shoham. 2020. The cost of training nlp models: A concise overview. *arXiv preprint arXiv:2004.08900*.
- Karishma Sharma, Yizhou Zhang, and Yan Liu. 2022. Covid-19 vaccine misinformation campaigns and social media narratives. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 920–931.
- Boris Shor and Nolan McCarty. 2011. The ideological mapping of american legislatures. *American Political Science Review*, 105(3):530–551.
- Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1):22–36.
- Sarah Shugars, Adina Gitomer, Stefan McCabe, Ryan J Gallagher, Kenneth Joseph, Nir Grinberg, Larissa Doroshenko, Brooke Foucault Welles, and David Lazer. 2021. Pandemics, protests, and publics: Demographic activity and engagement on twitter in 2020. *Journal of Quantitative Description: Digital Media*, 1.
- Denys Silveira, André Carvalho, Marco Cristo, and Marie-Francine Moens. 2018. **Topic Modeling using Variational Auto-Encoders with Gumbel-Softmax and Logistic-Normal Mixture Distributions**. In *International Joint Conference on Neural Networks (IJCNN)*.
- Yanchuan Sim, Bryan R Routledge, and Noah A Smith. 2015. The utility of text: The case of amicus briefs and the supreme court. In *Twenty-Ninth AAAI conference on artificial intelligence*.
- Barea Sinno, Bernardo Oviedo, Katherine Atwell, Malihe Alikhani, and Junyi Jessy Li. 2022. **Political ideology and polarization: A multi-dimensional approach**. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 231–243, Seattle, United States. Association for Computational Linguistics.
- Steven Sloman. 2005. *Causal models: How people think about the world and its alternatives*. Oxford University Press.
- Steven A Sloman and David Lagnado. 2015. Causality in thought. *Annual review of psychology*, 66:223–247.

- Aaron Smith, Christian Hardmeier, and Joerg Tiedemann. 2016. **Climbing mont BLEU: The strange world of reachable high-BLEU translations**. In *Proceedings of the 19th Annual Conference of the European Association for Machine Translation*, pages 269–281.
- Richard A Smith. 1995. Interest group influence in the us congress. *Legislative Studies Quarterly*, pages 89–139.
- Sarah A Smith and Annelise Russell. 2022. Different chambers, divergent rhetoric: Institutional differences and policy representation on social media. *American Politics Research*, page 1532673X221113017.
- Bob Sommer and John R Maycroft. 2008. Influencing public policy: An analysis of published op-eds by academics. *Politics & Policy*, 36(4):586–613.
- Frank Joseph Sorauf. 1988. *Money in American elections*. Glenview, Ill.: Scott, Foresman/Little, Brown College Division.
- Akash Srivastava and Charles Sutton. 2017. **Autoencoding variational inference for topic models**. In *Proceedings of the International Conference on Learning Representations*.
- Nisan Stiennon, Long Ouyang, Jeff Wu, Daniel M. Ziegler, Ryan J. Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul Christiano. 2020. **Learning to summarize from human feedback**. In *Proceedings of Advances in Neural Information Processing Systems*. Curran Associates, Inc.
- Petre Stoica and Yngve Selen. 2004. Model-order selection: a review of information criterion rules. *IEEE Signal Processing Magazine*, 21(4):36–47.
- Marilyn Strathern. 1997. **Improving Ratings: Audit in the british university system**. In *European Review*. Cambridge University Press.
- Thomas Stratmann. 1991. What do campaign contributions buy? deciphering causal effects of money and votes. *Southern Economic Journal*, pages 606–620.
- Thomas Stratmann. 2005. Some talk: Money in politics. a (partial) review of the literature. *Public Choice*, 124(1-2):135–156.
- Thomas Stratmann. 2009. How prices matter in politics: The returns to campaign advertising. *Public Choice*, 140:357–377.
- Haipeng Sun, Rui Wang, Kehai Chen, Masao Utiyama, Eiichiro Sumita, and Tiejun Zhao. 2020. **Knowledge distillation for multilingual unsupervised neural machine translation**. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3525–3535, Online. Association for Computational Linguistics.

- Naomi Sunderland, Tara Catalano, and Elizabeth Kendall. 2009. Missing discourses: concepts of joy and happiness in disability. *Disability & Society*, 24(6):703–714.
- Cass R Sunstein. 2018. *# Republic: Divided democracy in the age of social media*. Princeton University Press.
- Wendy K Tam Cho, James G Gimpel, and Iris S Hui. 2013. Voter migration and the geographic sorting of the american electorate. *Annals of the Association of American Geographers*, 103(4):856–870.
- Liling Tan, Jon Dehdari, and Josef van Genabith. 2015. **An awkward disparity between BLEU / RIBES scores and human judgements in machine translation**. In *Proceedings of the 2nd Workshop on Asian Translation (WAT2015)*, pages 74–81, Kyoto, Japan. Workshop on Asian Translation.
- Jiaxi Tang, Rakesh Shivanna, Zhe Zhao, Dong Lin, Anima Singh, Ed H. Chi, and Sagar Jain. 2020. **Understanding and Improving Knowledge Distillation**. *arXiv:2002.03532 [cs, stat]*.
- Raphael Tang, Yao Lu, and Jimmy Lin. 2019a. **Natural Language Generation for Effective Knowledge Distillation**. In *Proceedings of the 2nd Workshop on Deep Learning Approaches for Low-Resource NLP (DeepLo 2019)*, pages 202–208, Hong Kong, China. Association for Computational Linguistics.
- Raphael Tang, Yao Lu, Linqing Liu, Lili Mou, Olga Vechtomova, and Jimmy Lin. 2019b. **Distilling Task-Specific Knowledge from BERT into Simple Neural Networks**. *arXiv:1903.12136 [cs]*.
- Chris Tausanovitch and Christopher Warshaw. 2013. Measuring constituent policy preferences in congress, state legislatures, and cities. *The Journal of Politics*, 75(2):330–342.
- Silvia Terragni, Elisabetta Fersini, Bruno Giovanni Galuzzi, Pietro Tropeano, and Antonio Candelieri. 2021. **OCTIS: Comparing and optimizing topic models is simple!** In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 263–270, Online. Association for Computational Linguistics.
- Laure Thompson and David Mimno. 2020. Topic modeling with contextualized word representation clusters. *arXiv preprint arXiv:2010.12626*.
- Danielle M Thomsen and Bailey K Sanders. 2020. Gender differences in legislator responsiveness. *Perspectives on Politics*, 18(4):1017–1030.
- Runzhi Tian, Yongyi Mao, and Richong Zhang. 2020. **Learning VAE-LDA models with rounded reparameterization trick**. In *Proceedings of Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

- Ilya Tolstikhin, Sylvain Gelly, Olivier Bousquet, and Bernhard Scholkopf. 2018. Wasserstein Auto-Encoders. *ICLR*, page 16.
- Ciara Torres-Spelliscy. 2017. Time suck: How the fundraising treadmill diminishes effective governance. *Seton Hall Legis. J.*, 42:271.
- Shawn Treier. 2010. Where does the president stand? measuring presidential ideology. *Political Analysis*, 18(1):124–136.
- Shawn Treier and D Sunshine Hillygus. 2009. The nature of political ideology in the contemporary electorate. *Public Opinion Quarterly*, 73(4):679–703.
- Yariv Tsfati, Hajo G Boomgaarden, Jesper Strömbäck, Rens Vliegenthart, Alyt Damstra, and Elina Lindgren. 2020. Causes and consequences of mainstream media dissemination of fake news: literature review and synthesis. *Annals of the International Communication Association*, 44(2):157–173.
- Joshua A Tucker, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. 2018. Social media, political polarization, and political disinformation: A review of the scientific literature. *Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018)*.
- Keyon Vafa, Suresh Naidu, and David Blei. 2020. **Text-based ideal points**. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5345–5357, Online. Association for Computational Linguistics.
- Sebastián Vallejo Vera. 2023. By invitation only: on why do politicians bring interest groups into committees. *The Journal of Legislative Studies*, 29(1):1–38.
- Sebastián Vallejo Vera and Analía Gómez Vidal. 2022. The politics of interruptions: Gendered disruptions of legislative speeches. *The Journal of Politics*, 84(3):000–000.
- Wouter van Atteveldt and Tai-Quan Peng. 2018. When communication meets computation: Opportunities, challenges, and pitfalls in computational communication science. *Communication Methods and Measures*, 12(2-3):81–92.
- Robert Parks Van Houweling. 2003. *Legislators' personal policy preferences and partisan legislative organization*. Harvard University.
- Patrick Van Kessel, R Widjaya, S Shah, A Smith, and A Hughes. 2020. Congress soars to new heights on social media. *Pew Research Center, Washington, DC, July 16th*.
- Peter M VanDoren. 1990. Can we learn the causes of congressional decisions from roll-call data? *Legislative Studies Quarterly*, pages 311–340.

- Chris J Vargo, Lei Guo, and Michelle A Amazeen. 2018. The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New media & society*, 20(5):2028–2049.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.
- Jennifer Nicoll Victor and Gregory Koger. 2016. Financing friends: How lobbyists create a web of relationships among members of congress. *Interest Groups & Advocacy*, 5(3):224–262.
- Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *science*, 359(6380):1146–1151.
- Hanna M. Wallach, David M. Mimno, and Andrew McCallum. 2009. Rethinking LDA: Why Priors Matter. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems* 22, pages 1973–1981. Curran Associates, Inc.
- Rui Wang, Xuemeng Hu, Deyu Zhou, Yulan He, Yuxuan Xiong, Chenchen Ye, and Haiyang Xu. 2020a. Neural Topic Modeling with Bidirectional Adversarial Training. *ACL*, page 11.
- Rui Wang, Deyu Zhou, and Yulan He. 2020b. ATM: Adversarial-neural topic model. In *Proceedings of the Association for Computational Linguistics*.
- William Yang Wang. 2017. “liar, liar pants on fire”: A new benchmark dataset for fake news detection. pages 422–426.
- Duncan J Watts, David M Rothschild, and Markus Mobius. 2021. Measuring the news and its impact on democracy. *Proceedings of the National Academy of Sciences*, 118(15):e1912443118.
- Brian E Weeks and Homero Gil de Zúñiga. 2021. What’s next? six observations for the future of political misinformation research. *American Behavioral Scientist*, 65(2):277–289.
- Bernard L Welch. 1947. The generalization of ‘student’s’ problem when several different population variances are involved. *Biometrika*, 34(1-2):28–35.
- Frank Wilcoxon. 1992. Individual comparisons by ranking methods. In *Breakthroughs in statistics*, pages 196–202. Springer.
- John Wilkerson and Andreu Casas. 2017. Large-scale computerized text analysis in political science: Opportunities and challenges. *Annual Review of Political Science*, 20(1):529–544.

- Daniel Williams. 2022. The marketplace of rationalizations. *Economics & Philosophy*, pages 1–25.
- Jason Wilson. 2020. Disinformation and blame: How america’s far right is capitalizing on coronavirus. *The Guardian*, 19.
- Christopher Witko. 2006. Pacs, issue context, and congressional decisionmaking. *Political Research Quarterly*, 59(2):283–295.
- Christopher Witko, Jana Morgan, Nathan J Kelly, and Peter K Enns. 2021. *Hijacking the agenda: Economic power and political influence*. Russell Sage Foundation.
- Stefan Wojcik. 2018. Do birds of a feather vote together, or is it peer influence? *Political Research Quarterly*, 71(1):75–87.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. **Transformers: State-of-the-art natural language processing**. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.
- Andrew KC Wong and Manlai You. 1985. Entropy and distance of random graphs with application to structural pattern recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (5):599–609.
- Robert F Woolson. 2007. Wilcoxon signed-rank test. *Wiley encyclopedia of clinical trials*, pages 1–3.
- John R Wright. 1990. Contributions, lobbying, and committee voting in the us house of representatives. *American Political Science Review*, 84(2):417–438.
- Jiemin Wu, Yanghui Rao, Zusheng Zhang, Haoran Xie, Qing Li, Fu Lee Wang, and Ziyue Chen. 2020a. **Neural mixed counting models for dispersed topic discovery**. In *Proceedings of the Association for Computational Linguistics*. Association for Computational Linguistics.
- Xiaobao Wu, Chunping Li, Yan Zhu, and Yishu Miao. 2020b. **Short text topic modeling with topic distribution quantization and negative sampling decoder**. In *Proceedings of Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

- Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Łukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2016. **Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation**. *arXiv:1609.08144 [cs]*.
- Hongteng Xu, Wenlin Wang, Wei Liu, and Lawrence Carin. 2018. Distilled Wasserstein Learning for Word Embedding and Topic Modeling. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 1716–1725. Curran Associates, Inc.
- Guangxu Xun, Yaliang Li, Wayne Xin Zhao, Jing Gao, and Aidong Zhang. 2017. A correlated topic model using word embeddings. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence, IJCAI’17*, pages 4207–4213, Melbourne, Australia. AAAI Press.
- Liang Yang, Fan Wu, Junhua Gu, Chuan Wang, Xiaochun Cao, Di Jin, and Yuanfang Guo. 2020a. **Graph attention topic modeling network**. In *Proceedings of the World Wide Web Conference*. ACM.
- Weiwei Yang, Jordan Boyd-Graber, and Philip Resnik. 2017. Adapting topic models using lexical associations with tree priors. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1901–1906.
- Ziqing Yang, Yiming Cui, Zhipeng Chen, Wanxiang Che, Ting Liu, Shijin Wang, and Guoping Hu. 2020b. **TextBrewer: An Open-Source Knowledge Distillation Toolkit for Natural Language Processing**. *ACL Demo Session*.
- Tae Yano, Philip Resnik, and Noah A. Smith. 2010. **Shedding (a thousand points of) light on biased language**. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk*, pages 152–158, Los Angeles. Association for Computational Linguistics.
- Moran Yarchi, Christian Baden, and Neta Kligler-Vilenchik. 2021. Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication*, 38(1-2):98–139.
- Hao Zhang, Bo Chen, Dandan Guo, and Mingyuan Zhou. 2018. **WHAI: weibull hybrid autoencoding inference for deep topic modeling**. In *Proceedings of the International Conference on Learning Representations*.

- He Zhao, Lan Du, Wray L. Buntine, and Mingyuan Zhou. 2018. **Dirichlet belief networks for topic structure learning**. In *Proceedings of Advances in Neural Information Processing Systems*. Curran Associates, Inc.
- He Zhao, Dinh Phung, Viet Huynh, Yuan Jin, Lan Du, and Wray Buntine. 2021a. **Topic modelling meets deep neural networks: A survey**. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 4713–4720. International Joint Conferences on Artificial Intelligence Organization. Survey Track.
- He Zhao, Dinh Phung, Viet Huynh, Trung Le, and Wray Buntine. 2021b. **Neural topic model via optimal transport**. In *Proceedings of the International Conference on Learning Representations*.
- Ruiqi Zhong, Charlie Snell, Dan Klein, and Jacob Steinhardt. 2022. Describing differences between text distributions with natural language. In *International Conference on Machine Learning*, pages 27099–27116. PMLR.
- Deyu Zhou, Xuemeng Hu, and Rui Wang. 2020. **Neural topic modeling by incorporating document relationship graph**. In *Proceedings of Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Qile Zhu, Zheng Feng, and Xiaolin Li. 2018. **GraphBTM: Graph enhanced autoencoded variational inference for biterm topic model**. In *Proceedings of Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Melissa Zimdars. 2016. My “fake news list” went viral. but made-up stories are only part of the problem. *Washington Post*.
- George K. Zipf. 1949. *Human Behaviour and the Principle of Least Effort*. Addison-Wesley.