What Sounds "Right" to Me? Experiential Factors in the Perception of Political Ideology

Qinlan Shen
Carnegie Mellon University
qinlans@cs.cmu.edu

Carolyn P. Rosé
Carnegie Mellon University
cprose@cs.cmu.edu

Abstract

In this paper, we challenge the assumption that political ideology is inherently built into text by presenting an investigation into the impact of experiential factors on annotator perceptions of political ideology. We construct an annotated corpus of U.S. political discussion, where in addition to ideology labels for texts, annotators provide information about their political affiliation, exposure to political news, and familiarity with the source domain of discussion, Reddit. We investigate the variability in ideology judgments across annotators, finding evidence that these experiential factors may influence the consistency of how political ideologies are perceived. Finally, we present evidence that understanding how humans perceive and interpret ideology from texts remains a challenging task for state-of-the-art language models, pointing towards potential issues when modeling user experiences that may require more contextual knowledge.

1 Introduction

Social media companies, like Twitter, Facebook, and Reddit, play an important role in political discourse by providing a space for users to interact with different viewpoints. Understanding political discussion on these platforms often requires one to identify the ideologies behind texts, as understanding the viewpoints reflected in a text can provide insight into the partisanship of beliefs (Monroe et al., 2008) or the persuasive strategies used by different ideological groups (Tsur et al., 2015).

Prior research on political discussion often relies on a "ground-truth" to aid in obtaining ideology labels for social media data. For example, due to the scale of political content on social media, a common paradigm is to obtain some ground-truth labels that are propagated to a larger set of texts using semi-supervised learning (Lin and Cohen, 2010; Zhou et al., 2011). The relationship between a social media artifact and various forms of established political knowledge can also be used to ground or validate ideology labels. Some examples of this include using author interactions with politicians with known party affiliations (Djemili et al., 2014; Barberá, 2015), ideological communities (Chandrasekharan et al., 2017; Shen and Rosé, 2019), and central users (Pennacchiotti and Popescu, 2011; Preotiuc-Pietro et al., 2017) as a starting heuristic, or evaluating a labeling approach by comparing geolocation tags attached to posts with historical voting patterns (Demszky et al., 2019).

A limitation of these approaches, however, is that behavior on social media does not evenly or uniformly reflect the held political beliefs of participants. While there is evidence that people tend to engage with others who share similar beliefs (Halberstam and Knight, 2016), people also commonly interact with or even seek out communities and users they do not agree with (Kelly et al., 2005; Tan et al., 2016). Additionally, the practice of displaying one's political beliefs, which many grounding techniques rely on, varies in prevalence across online communities (Lampe et al., 2007; Zhong et al., 2017; Pathak, 2020). The concept of linguistic agency (Goffman et al., 1978) also challenges the idea that individual factors, such as ideology, are predictably presented in text. Based on an author's social goals for participating in political discussion, it may not be contextually relevant to project a strong impression of their political ideology. People engaged in interactive political discussion, however, still form perceptions about the alignments of others based on how they sound, often relying on their own conceptions of ideology in the process.

The issue of perceiving ideology also plays a role when ideology labels are obtained using crowdsourced annotators. While making judgments, the

Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics, pages 1762–1771
April 19 - 23, 2021.

annotator plays a similar role to a user participating in the discussion when perceiving the ideology of the speaker behind a text. However, annotators are expected to assign an explicit ideology label to a text with less contextual knowledge about how the text was produced. Thus, annotators may rely heavily on their own *experiential* factors, such as one's own beliefs or level of political engagement, when considering ideology. As a result, this process may introduce inconsistencies and biases in ideological labels used for political analysis.

In this paper, we present an exploration of how experiential factors play a role in how annotators perceive ideology in text. Building upon prior work investigating annotation bias (Zaidan and Callison-Burch, 2011; Waseem, 2016; Joseph et al., 2017; Ross et al., 2017; Schwartz et al., 2017; Geva et al., 2019), we construct an annotated corpus of posts from political subcommunities on Reddit but incorporate additional contextual information about the annotators making ideology judgments.[1] While previous work (Joseph et al., 2017) has shown that source-side contextual features, such as user profiles and previous tweets, can influence label quality in stance annotation, we focus our analyses on contextual factors on the side of annotators. Most similar to our work, Carpenter et al. (2017) and Carpenter et al. (2018) examine the impact of an annotator's identity and openness on their ability to accurately assess author attributes, including political orientation. In our work, however, we examine the impact of an annotator's political beliefs, knowledge, and Reddit familiarity, on their judgments, using factors more specific to political participation on Reddit. We additionally consider the issue of annotator bias in ideology labeling not as an issue of accuracy but rather an issue of social variability. Under this view, we evaluate the performance of a state-of-the-art language model on its capacity to mirror different human perceptions of ideology to examine whether extralinguistic factors introduced through annotation may degrade model performance compared to other labels.

2 Dataset Construction

Our dataset is drawn from the popular content aggregation and discussion platform Reddit. Political discussion on Reddit is centered on *subreddits*, subcommunities centered on support for specific political candidates, organizations, and issues. For our analyses, we aim to label political distinctions on Reddit along the left-right political spectrum in U.S. politics. Using the monthly dumps from May to September 2019 from the Reddit Pushshift API (Baumgartner et al., 2020), we collect all submissions and comments from the top political subreddits[2] by subscriber count. The collected subreddits were manually labeled as left or right, based on the subreddit description and top posts. We then select the top 12 left and top 12 right subreddits from the monthly dumps where discussion is primarily focused on U.S. politics.[3] The selected subreddits are shown in Table 3 (Supplementary Material).

2.1 Paired Ideology Ranking Task

Prior work on annotating viewpoints (Iyyer et al., 2014; Bamman and Smith, 2015) generally presents annotators with texts in isolation to label with an ideology of interest. One drawback of this approach is the high degree of political expertise annotators are required to have to recognize that a text matches an ideology. To reduce the amount of overhead in recruiting and training political annotators, we present annotators instead with a paired ideology ranking task. Rather than examining texts in isolation, annotators are shown two texts and asked to select the text that is more likely to be authored by someone with the ideology of interest.

For our setup, our goal is to pair a text authored by a left-leaning user with one by a right-leaning user. We use a heuristic-based semi-supervised approach to label texts based on the subreddit participation patterns of their authors. To expand the set of subreddits with ideological labels, we label all subreddits in the monthly dump data as left, neutral, or right based on user overlap with the 24 political subreddits with a known ideological slant (Section 2). For each subreddit, we calculate the z-score of the log odds ratio of a user participating in that subreddit and a known left-leaning subreddit vs. a right-leaning subreddit. A subreddit is labeled as either "left" or "right" if the calculated z-score satisfies a one-tailed Z test at $p = 0.05$ in the corresponding direction or "neutral" otherwise. Authors are then labeled based on their distribution of participation on the left vs. right subreddits. While

[1] This study was approved by the institutional review board at our institution.

[2] https://www.reddit.com/r/redditlists/comments/josdr/list_of_political_subreddits/

[3] r/politics was not included due to its initial history as a default subreddit contributing to its high subscriber count.

users on Reddit have been shown to primarily engage with pro-social home communities (Datta and Adar, 2019) and similar heuristics have been used in prior work as an indicator of user interests and/or ideology (Olson and Neal, 2015; Chandrasekharan et al., 2017; Shen and Rosé, 2019), we emphasize that we use this heuristic to create a basis of comparison, rather than assuming that it provides "correct" ideology labels.

In order to ensure that the text comparison helps annotators to perceive ideological differences, rather than presenting two unrelated texts that are essentially considered in isolation, we want to present paired texts that are similar in content. As a first step for generating comparisons with similar content, we require paired texts to discuss the same entity, since political discussions are primarily centered on the politicians, organizations, and geopolitical entities influencing policy decisions. To identify entities of interest, we use Stanford Core NLP (Manning et al., 2014) to extract occurrences of people, locations, organizations, and ideologies over our corpus of 24 subreddits. We limit entities under consideration to those that have occurred at least 300 times in our corpus and are easy to disambiguate. The considered entities are shown in Table 4 (Supplementary Material).

To limit the impact of confounds, such as topic or entity salience, when comparing texts with the same entity, we use propensity score matching (Rosenbaum and Rubin, 1983) to match each left-aligned text with a right-aligned text that discusses the same entity in a similar context. A subset of 65 pairs was manually curated to use as screening questions to ensure that workers had a baseline knowledge of U.S. politics. These screening pairs were selected to be easier than the main task pairs – they are more limited in which entities discussed and express more explicit and/or extreme attitudes.

2.2 Annotation Task Details

We recruit workers on Amazon Mechanical Turk to complete our paired ideological ranking task. Given a pair of texts discussing the same highlighted political entity, we ask annotators to determine which of the two posts is more likely to have been written by someone who is either left-leaning or right-leaning. Annotators were instructed to use as many contextual cues as possible to form an impression of the political views held by the authors of the texts. To provide some guidance to annotators for what cues to consider, we train workers to consider the following features in the instructions:

- **Attitude:** evaluation in favor of or against an entity. Ex: *I trust Bernie* from someone who favors Bernie Sanders (left).
- **Positioning:** situating one's viewpoint with respect to the entity's. Ex: *Listen to the Dems* refers to Democrats as an out-group (right).
- **Jargon:** use of speciality in-group vocab. Ex: *Trump GEOTUS!* – "God-Emperor" abbreviation specific to Trump supporters (right).

The annotation task is shown in Figure 1 (Supplementary Material). Each worker was asked to annotate 18 pairs from our main task set and 8 screening questions, which were scattered throughout the assignment as an attention check. For each main task pair, we assign up to 5 workers for annotation. We restrict the worker pool to the U.S. and filter out workers who scored less than a 75% on the screening questions. Overall, we collect annotations for 630 non-screening pairs.

2.3 Annotator Background Post-Survey

After the annotation task, workers were asked to complete a survey (questions listed in Supplementary Material A) to assess their political affiliation, exposure to U.S. political news, and familiarity with political discussion on Reddit. Answers to the survey were inspected manually to assign annotators labels along three identifier categories:

- **Political ideology:** This category indicates the annotator's political ideology. Annotators are labeled as *left*, *center*, or *right* based on their self-identified ideology and affiliation with U.S. political parties.
- **News access:** This category indicates the annotator's exposure to political news. Annotators are labeled as *news* or *non-news* based on how frequently they access news on the 2020 U.S. presidential election.
- **Reddit familiarity:** This category indicates the annotator's familiarity with participation in political discussion on Reddit. Annotators are labeled as a *redditor* or a *non-redditor* based on their level of participation on Reddit in the past year. Redditors are further subdivided into *political* and *non-political* redditors based on their familiarity with the political subreddits included in our corpus.

		# workers	α
Overall	-	158	0.388
Ideology	left	89	0.427
	right	43	0.372
	center	26	0.325
News	news	126	0.393
	non-news	32	0.336
Reddit	redditor	114	0.393
	–political	86	0.389
	–non-political	28	0.430
	non-redditor	44	0.359

Table 1: Number of workers and Krippendorff's α agreement within the annotator groups over the full non-screening set. Agreement over other question sets can be found in Table 5 (Supplementary Material)

3 Dataset Statistics and Analysis

3.1 Annotator Demographics

Of the 180 recruited workers initially recruited for the task, 22 were discarded for answering fewer than 75% of the screening questions correctly, giving us a final pool of 158 annotators. Table 1 illustrates the distribution of the remaining workers across labels within the three categories. Labels across categories do not appear to be correlated (mean variance inflation factor = 1.043).

3.2 Agreement/Difference Results

We use Krippendorf's α (Krippendorff, 2004) to evaluate annotator agreement on our task to account for different user pools for each question. Despite a high degree of agreement across the pool of screening questions ($\alpha = 0.7311$), the overall agreement across annotators in our general, non-screening set is relatively low ($\alpha = 0.3878$), suggesting that the task of predicting the ideology of a text is nuanced and open to interpretation.

We also calculate agreement for workers within each of our annotator groups (Table 1) in order to examine whether annotators with similar backgrounds are more likely to perceive ideology similarly. Overall, in-group agreement remains around the same level as the general task. However, an interesting pattern across annotator labels is that workers who are less likely to be familiar with the expression of political ideology on Reddit – non-redditors ($\alpha = 0.359$), people who do not frequently read political news ($\alpha = 0.336$), and people who do not identify with the left or right ($\alpha = 0.325$) – have lower agreement. This suggests that familiarity with the norms of political discussion on Reddit may contribute to a more consistent perception of ideology for Reddit texts.

We additionally use McNemar's chi-squared test over pairwise comparisons of annotator groups under the same category to examine whether annotators with different backgrounds differ in their judgments. To ground the comparison, we evaluate annotator groups based on whether the majority of workers in the group gave the same answer as our semi-supervised labels (Section 2.1). Because these semi-supervised labels only provide a noisy estimate of ideology, we use these labels to create a basis of comparison. Rather than to check how "accurately" each group estimates ideology, this heuristic allows us to specifically quantify differences in judgments between groups. We find that for all comparison pairs, groups differ significantly in their answers over the same questions. In our pairwise comparisons, we also saw that the ideology of the annotator contributes heavily to variability in annotator judgments. The two groups with the highest percentage of questions with mismatched answers are left-leaning and right-leaning annotators, and 3 of the top 4 comparison pairs with the most mismatched answers are between ideology groups (Supplementary Material Table 6).

3.3 Sources of Variability

To examine possible explanations for the variability in annotator judgments across groups, we focus primarily on differences in judgments between left-leaning and right-leaning annotators. When examining differences at the entity-level, we find that the entities with the most mismatches tended to be highly visible entities that had a strong connection to a particular party during the 2020 election, such as highly visible political figures (e.g. Joe Biden, Nancy Pelosi) or the most common ideologies associated with each side (e.g. Republican Party, conservatism, liberalism), compared to less salient entities. This is unsurprising, as we expect people to develop different conceptions of salient entities building up to major events like elections, even with relatively limited media exposure.

Finally, to investigate what aspects of the posts themselves contributed to variations in judgments between left-leaning and right-leaning workers, we ran a salience (Monroe et al., 2008) analysis for

mismatched question pairs with highly visible entities. We found that annotators were less likely to select a post that expresses explicit abuse towards an opposing entity as being authored by someone with the same political views as themselves. For example, a right-leaning annotator was less likely to consider a post calling Biden a "pedophile" as right-leaning compared to liberal annotator. This may suggest that social desirability bias (Krumpal, 2013), may have an impact on decision-making, even when the task is not directly related to collecting data about the annotator themselves.

4 Perceptions vs. Heuristic Labels

Prior work (Castelle, 2018) suggests that deep text classification models perform poorly when labels are influenced by extralinguistic contextual factors. While the semi-supervised labels that we generated are based on a behavioral heuristic outside of the text, our analyses of human judgments suggest that the annotation process introduced additional interactional factors into ideological labeling. We investigate whether these factors influence model performance by evaluating a BERT-based (Devlin et al., 2019) model on its ability to match human judgments on the paired ideology ranking task.

For our evaluation model, we finetune BERT-mask on the 24 subreddit corpus. Next, for each text, we average its contextual embeddings in two ways: over (a) all tokens in the text and (b) all entity-related tokens in the text. We then concatenate the averaged embeddings, then use the resulting vector as input to a pairwise logistic regression model. For each annotator group, we use the majority answer for each question as the group label.

Table 2 shows the performance of the model on the full 630 pair non-screening set. For all annotator groups, we found that the model has a significant drop in performance when asked to match human judgments vs. labels generated through our semi-supervised heuristic on the same dataset. To examine whether this drop in performance was due to inconsistencies in human judgments on particularly difficult or contentious distinctions, we additionally present results on a higher consensus subset ($\alpha = 0.6216$) of 459 text pairs, where at least 75% of workers select the same answer. We found that while there was a small increase in performance on matching human judgments on the high consensus subset for all groups, performance still dropped compared to the semi-supervised labels, suggesting that matching human understanding of ideology is challenging for these models.

	SS (F)	**H (F)**	**SS (C)**	**H (C)**
Overall	69.28*	56.82	70.16*	58.62
left	68.60*	60.13	69.75*	63.00
right	65.97*	53.41	69.51*	56.49
center	63.81*	52.58	70.41*	60.27
news	67.48*	57.43	68.86*	59.43
non-news	66.01*	54.03	67.89	63.27
redditor	70.89*	57.65	69.46*	59.95
–political	70.59*	57.25	67.90*	59.01
–non-political	69.62*	57.01	72.63*	58.57
non-redditor	65.89*	51.65	64.06*	55.49

Table 2: F1 scores for a BERT-based ranking model on semi-supervised (**SS**) and human annotator (**H**) labels for the full non-screening set (**F**) and a high-consensus subset (**C**). $*p < 0.05$ difference in performance between the semi-supervised and human annotator labels

5 Conclusion and Future Work

In this paper, we reconsider the idea of ground-truth labels of political ideology and investigate the impact of experiential factors on human perception of ideology in text. We construct and analyze an annotated corpus that incorporates experiential information about annotators, finding evidence that annotator backgrounds influence the consistency of political ideology judgments and that current classification models struggle to match human perceptions of ideology across different groups. From our analyses on factors contributing to variations in judgments, there is a greater need for targeted recruiting of annotators that are familiar with and contextualized to the domain being annotated. In future work, we aim to extend our investigation to examine how stylistic elements of text contribute to people's perception of political ideologies in interaction. These analyses may provide further insight into the effectiveness of political communication strategies or the differences in how political groups interact with in-group and out-group members.

Acknowledgements

This work was supported in part by NSF Grant IIS 1546393 and the K&L Gates Presidential Fellowship.

References

David Bamman and Noah A Smith. 2015. Open extraction of fine-grained political statements. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 76–85.

Pablo Barberá. 2015. Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis*, 23(1):76–91.

Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The pushshift reddit dataset. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 830–839.

Jordan Carpenter, Daniel Preotiuc-Pietro, Jenna Clark, Lucie Flekova, Laura Smith, Margaret L Kern, Anneke Buffone, Lyle Ungar, and Martin Seligman. 2018. The impact of actively open-minded thinking on social media communication. *Judgment and Decision Making*, 13(6):562.

Jordan Carpenter, Daniel Preotiuc-Pietro, Lucie Flekova, Salvatore Giorgi, Courtney Hagan, Margaret L Kern, Anneke EK Buffone, Lyle Ungar, and Martin EP Seligman. 2017. Real men don't say "cute" using automatic language analysis to isolate inaccurate aspects of stereotypes. *Social Psychological and Personality Science*, 8(3):310–322.

Michael Castelle. 2018. The linguistic ideologies of deep abusive language classification. In *Proceedings of the 2nd Workshop on Abusive Language Online (ALW2)*, pages 160–170.

Eshwar Chandrasekharan, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. 2017. You can't stay here: The efficacy of reddit's 2015 ban examined through hate speech. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):1–22.

Srayan Datta and Eytan Adar. 2019. Extracting inter-community conflicts in reddit. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, pages 146–157.

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 Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 2970–3005.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics*.

Sarah Djemili, Julien Longhi, Claudia Marinica, Dimitris Kotzinos, and Georges-Elia Sarfati. 2014. What does Twitter have to say about ideology? In *NLP 4 CMC: Natural Language Processing for Computer-Mediated Communication/Social Media-Pre-conference workshop at Konvens 2014*, volume 1. Universitätsverlag Hildesheim.

Mor Geva, Yoav Goldberg, and Jonathan Berant. 2019. Are We Modeling the Task or the Annotator? An Investigation of Annotator Bias in Natural Language Understanding Datasets. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1161–1166.

Erving Goffman et al. 1978. *The Presentation of Self in Everyday Life*. Harmondsworth London.

Yosh Halberstam and Brian Knight. 2016. Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter. *Journal of Public Economics*, 143:73–88.

Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. 2014. Political ideology detection using recursive neural networks. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1113–1122.

Kenneth Joseph, Lisa Friedland, William Hobbs, David Lazer, and Oren Tsur. 2017. ConStance: Modeling Annotation Contexts to Improve Stance Classification. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1115–1124.

John Kelly, Danyel Fisher, and Marc Smith. 2005. Debate, division, and diversity: Political discourse networks in USENET newsgroups. In *Online Deliberation Conference*, pages 1–35. Stanford University.

Klaus Krippendorff. 2004. Reliability in content analysis: Some common misconceptions and recommendations. *Human communication research*, 30(3):411–433.

Ivar Krumpal. 2013. Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, 47(4):2025–2047.

Cliff AC Lampe, Nicole Ellison, and Charles Steinfield. 2007. A familiar face (book) profile elements as signals in an online social network. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 435–444.

Frank Lin and William W Cohen. 2010. Semi-supervised classification of network data using very few labels. In *2010 International Conference on Advances in Social Networks Analysis and Mining*, pages 192–199. IEEE.

Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60.

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.

Randal S Olson and Zachary P Neal. 2015. Navigating the massive world of reddit: Using backbone networks to map user interests in social media. *PeerJ Computer Science*, 1:e4.

Arjunil Pathak. 2020. *Extraction and Analysis of Self Identity in Twitter Biographies*. Ph.D. thesis, State University of New York at Buffalo.

Marco Pennacchiotti and Ana-Maria Popescu. 2011. Democrats, Republicans and Starbucks Afficionados: User Classification in Twitter. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 430–438.

Daniel Preotiuc-Pietro, Ye Liu, Daniel Hopkins, and Lyle Ungar. 2017. Beyond binary labels: political ideology prediction of twitter users. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 729–740.

Paul R Rosenbaum and Donald B Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Björn Ross, Michael Rist, Guillermo Carbonell, Benjamin Cabrera, Nils Kurowsky, and Michael Wojatzki. 2017. Measuring the reliability of hate speech annotations: The case of the european refugee crisis. *arXiv preprint arXiv:1701.08118*.

Roy Schwartz, Maarten Sap, Ioannis Konstas, Leila Zilles, Yejin Choi, and Noah A Smith. 2017. The Effect of Different Writing Tasks on Linguistic Style: A Case Study of the ROC Story Cloze Task. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 15–25.

Qinlan Shen and Carolyn Rosé. 2019. The Discourse of Online Content Moderation: Investigating Polarized User Responses to Changes in Reddit's Quarantine Policy. In *Proceedings of the Third Workshop on Abusive Language Online*, pages 58–69.

Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th International Conference on World Wide Web*, pages 613–624.

Oren Tsur, Dan Calacci, and David Lazer. 2015. A frame of mind: Using statistical models for detection of framing and agenda setting campaigns. 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 1629–1638.

Zeerak Waseem. 2016. Are You a Racist or Am I Seeing Things? Annotator Influence on Hate Speech Detection on Twitter. In *Proceedings of the First Workshop on NLP and Computational Social Science*, pages 138–142.

Omar Zaidan and Chris Callison-Burch. 2011. Crowdsourcing Translation: Professional Quality from Non-Professionals. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 1220–1229.

Changtao Zhong, Hau Wen Chang, Dmytro Karamshuk, Dongwon Lee, and Nishanth Sastry. 2017. Wearing many (social) hats: How different are your different social network personae? In *11th International Conference on Web and Social Media, ICWSM 2017*, pages 397–406. AAAI press.

Daniel Xiaodan Zhou, Paul Resnick, and Qiaozhu Mei. 2011. Classifying the political leaning of news articles and users from user votes. In *Proceedings of the International AAAI Conference on Web and Social Media*.

	Selected subreddits
Left	r/LateStageCapitalism, r/SandersForPresident, r/democrats, r/socialism, r/Liberal, r/VoteBlue, r/progressive, r/ChapoTrapHouse, r/neoliberal, r/esist, r/The_Mueller, r/The_Mueller
Right	r/The_Donald, r/Libertarian, r/Republican, r/Conservative, r/JordanPeterson, r/TheNewRight, r/Anarcho_Capitalism, r/conservatives, r/ShitPoliticsSays, r/POLITIC, r/AskTrumpSupporters, r/AskThe_Donald

Table 3: Selected subreddits included in the construction of the dataset and their ideological alignments.

	Selected entities
People	*Donald Trump*, *Joe Biden*, *Bernie Sanders*, *Barack Obama*, *Hillary Clinton*, *Robert Mueller*, *Nancy Pelosi*, *Kamala Harris*, *Alexandria Ocasio-Cortez*, Andrew Yang, Elizabeth Warren, Pete Buttigieg
Ideologies	*conservatives/conservatism*, *liberals/liberalism*, libertarians/libertarianism, socialists/socialism, capitalists/capitalism
Organizations	*Republican Party/Republicans*, *Democratic Party/Democrats*, Congress
Locations	Russia

Table 4: Selected entities included in the construction of the dataset. Italicized entities are also included in the screening set.

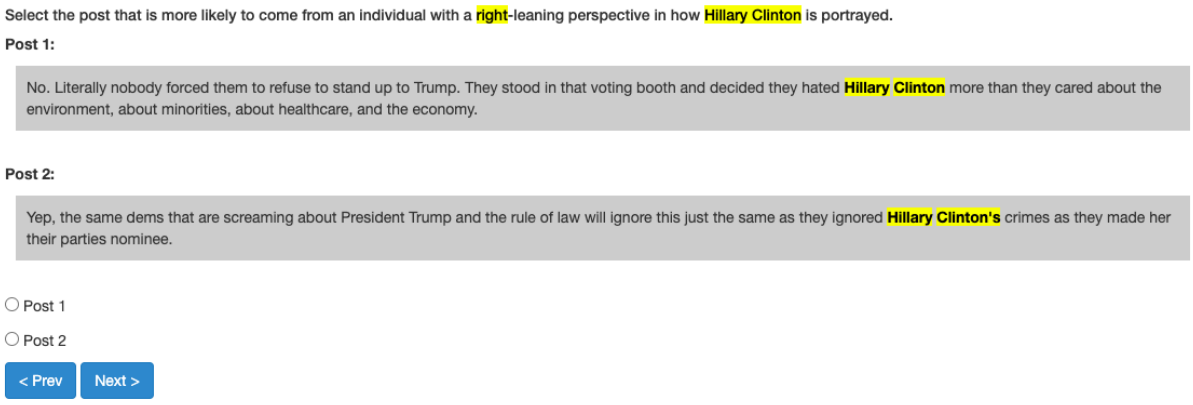

Figure 1: Screenshot of a question in the paired ideological annotation task. Annotators are presented with two texts discussing the same highlighted entity in a similar context, one from a left-leaning user and another from a right-leaning user based on a semi-supervised labeling heuristic. Annotators are asked to select which of the two texts is more likely to be authored by someone with the highlighted ideology.

A Survey Questions

A.1 Political ideology

1. Please indicate where you identify on the liberal-conservative spectrum.
    - Liberal
    - Somewhat liberal
    - Moderate
    - Somewhat conservative
    - Conservative
    - I don't know
2. Please indicate how strongly you identify with the following U.S. political parties.
    - Parties
        - Democratic Party
        - Republican Party
        - Libertarian Party
        - Green Party
        - Constitution Party
        - Democratic Socialists of America
        - Reform Party
    - Responses
        - I do not identify with this party
        - Somewhat identify
        - Identify
        - Strongly identify
        - I don't know

A.2 News access

1. On average, how often did you check the news related to the 2020 presidential election in the U.S. in the past year?
    - Never
    - Less than once a month
    - A few times a month
    - Once a week
    - Several times a week
    - Once a day
    - Several times a day

A.3 Reddit familiarity

1. On average, how often have you visited Reddit in the past year?
    - Never
    - Less than once a month
    - A few times a month
    - Once a week
    - Several times a week
    - Once a day
    - Several times a day
2. On average, how often have you posted content to Reddit in the past year?
    - Never
    - Less than once a month
    - A few times a month
    - Once a week
    - Several times a week
    - Once a day
    - Several times a day
3. Please indicate your familiarity with the following subreddits (listed in Table 3).
    - I have never heard of this subreddit
    - I have heard of but never accessed this subreddit
    - I have accessed or posted on this subreddit at least once
    - I sometimes access or post on this subreddit
    - I often access or post on this subreddit

B Detailed Agreement Results

		Agreement		
		F	**S**	**C**
Overall	-	0.388	0.731	0.621
Ideology	left	0.427	0.632	0.674
	right	0.372	0.638	0.583
	center	0.325	0.493	0.609
News	news	0.393	0.627	0.622
	non-news	0.336	0.638	0.588
Reddit	redditor	0.393	0.644	0.622
	–political	0.389	0.641	0.615
	–non-political	0.430	0.582	0.676
	non-redditor	0.359	0.586	0.608

Table 5: Krippendorff's α agreement results for survey categories for the full non-screening annotated set (F), the screening questions (S), and the high-consensus questions subset (C).

C Mismatch Statistics

Group comparison	% mismatch
left/right	28.15
right/center	26.97
non-political/non-redditor	26.47
left/center	24.44
right/non-news	23.74
non-political/right	23.17
news/non-news	22.71
non-political/political	22.61
non-political/center	21.05
non-redditor/political	20.40

Table 6: Comparison pairs with highest percentage of questions where the majority gave different answers.

Entity	% mismatch
libertarians/libertarianism	100.0
Republican Party/Republicans	53.85
Russia	43.75
conservatives/conservatism	42.86
Hillary Clinton	39.13
Joe Biden	38.89
Nancy Pelosi	36.36
liberals/liberalism	31.81
Robert Mueller	28.57
Alexandria Ocasio-Cortez	26.67

Table 7: Entities with highest percentage of questions where the left-leaning and right-leaning annotators gave different answers.

D Human Judgments vs. Labels

		Worker Match
Overall	-	68.53
Ideology	left	70.05
	right	67.30
	center	65.38
News	news	69.24
	non-news	65.80
Reddit	redditor	68.49
	–political	69.03
	–non-political	66.87
	non-redditor	68.65

Table 8: Average percentage of human judgments that match with semi-supervised labels per annotation group.