P1SCO: Social Dimensions from a Perspectivist Lens

Anonymous ACL submission

Abstract

We introduce P1SCO, a dataset of social media comments collected from three distinct platforms, annotated according to ten social dimensions to capture the diversity of social interactions and perceptions. The dataset is carefully disaggregated to allow analysis at the level of individual comment, annotator, and platform. In addition to social dimension labels, we include rich metadata on the annotators, including demographics, Big Five personality profiles, and political affiliation. This combination of comment-level annotations and annotator-level features enables nuanced analyses of how social perception varies across platforms, individual differences, and demographic factors. By preserving the diversity of annotator perspectives, our dataset supports studies of inter- and intra-annotator agreement, the influence of personality and political orientation on social interpretation, and the cross-platform dynamics of social discourse. We release the dataset to facilitate research in computational social science, social psychology, and natural language processing, promoting reproducibility and enabling further exploration of social dimension perception across heterogeneous online communities.

1 Introduction

In *Ways of Seeing*, John Berger argues that we do not simply look at things—we interpret them through layers of culture, ideology, and personal experience. What we "see" is already shaped by power, history, gender norms, class relations, and our own positionality. The same thing can be said about language: if pragmatics is about meaning in context, then every individual brings their own context, shaping interpretation in ways as varied and layered as the perspectives Berger describes (see Fig 1). Studying pragmatic expression and idiolects reveals not just how we speak, but how society itself is structured through interaction.

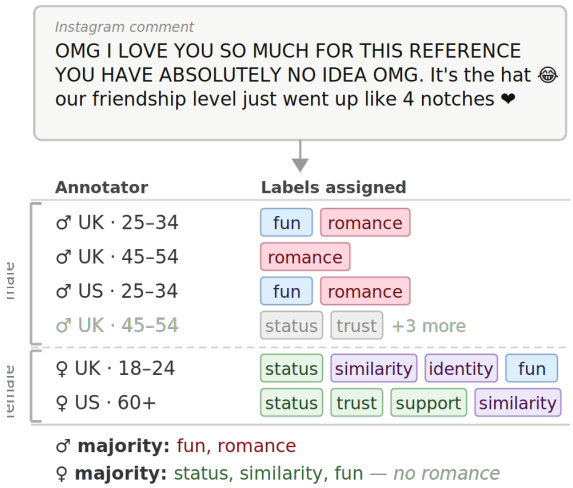


Figure 1: The same Instagram comment, interpreted through different demographic lenses. Male annotators assign romance and fun; female annotators assign status, similarity, and fun. Disagreement reflects genuine perspectival variation, not annotation error.

The ambiguous nature of language and the idiosyncrasies of human label variation have inspired a growing body of literature studying the factors that affect different people's understanding of an utterance (e.g., Ovesdotter Alm, 2011; Hovy and Yang, 2021; Plank, 2022). This work takes a perspectivist approach and acknowledges that linguistic expression is a performance of identity.

Previous work in computational social science has shown that conversations can be meaningfully analyzed along ten social dimensions that make up the fundamental building blocks of human relations (Choi et al., 2020). These dimensions capture the social intent underlying utterances in a way that both interlocutors can *perceive*. Crucially, while these dimensions provide a structured framework for understanding social intent, their interpretation is far from uniform. Each speaker brings a unique set of experiences, personality traits, and cultural background to an interaction, thus shaping how

they express—and how they perceive—the same social intent. This interplay between social dimensions and individual variation highlights the perspectivist nature of language: just as Berger observed that we see the world through layered lenses, we also enact and interpret social meaning through personal and situational filters. By examining conversations through both these lenses, we can uncover patterns in how social goals are communicated and understood across diverse individuals.

By treating both social intent and individual variation as central to understanding language, we can begin to model how context and perspective jointly shape meaning in conversation and human relationships. To this end, we introduce P1SCO (pronounced *pisco*), Perspectives on the 10 Social dimensions of Conversation, a large dataset of social media comments from diverse platforms, manually labelled by diverse annotators.[1]

2 Background and Related Work

Through an extensive review of the social science literature, prior work identified ten dimensions of the social pragmatics of language (Deri et al., 2018). These dimensions capture archetypal signals of social intent typically expressed in everyday conversations, such as conveying appreciation or offering emotional support. Table 1 reports the full set of dimensions along with their concise definitions. Beyond their linguistic realization, these dimensions also correspond to core mental categories that individuals typically refer to when conceptualizing their social relationships. As such, they provide a general theoretical framework for the study of social interaction across diverse contexts. Accordingly, the ten dimensions framework has been widely adopted to investigate several types of social dynamics including agreement (Monti et al., 2022), coordination (Lucchini et al., 2022; La Cava et al., 2023), and well-being (Aiello et al., 2021; Balsamo et al., 2023) in online communities, as well as to simulate human interaction dynamics with Large Language Models (Breum et al., 2024).

To operationalize these dimensions, prior work introduced a human-annotated corpus of social media text designed to train supervised natural language processing models for detecting signals of social pragmatics from conversations (Choi et al., 2020). The corpus consists of 7855 sentences sampled from Reddit comments selected at random. Each sentence contains between 6 and 20 words and includes at least one first- or second-person pronoun, to focus on phrases that follow a direct conversational structure. Each sentence was annotated by three to five Amazon Mechanical Turk workers, who were allowed to assign multiple labels when applicable. As the first dataset of its kind, this dataset is limited to a single online platform and is released in an aggregated form, reporting only total label counts per sentence and providing no information on annotator demographics.

Recent work in NLP has begun to argue that social meaning is inherently perspectival and that annotator disagreement may reflect legitimate interpretive variation rather than noise (Plank, 2022; Cabitza et al., 2023). Work in this area has thus focused on both sociodemographic (Pei and Jurgens, 2023; Hovy and Yang, 2021) and individual characteristics (Orlikowski et al., 2025) to enrich models of linguistic phenomena. A number of datasets have been published in recent years reflecting a perspectivist paradigm but these datasets are predominantly focused on offensiveness and sarcasm (Frenda et al., 2025).

Social media platforms differ substantially in demographics and interactional norms, leading to systematic variation in how content is produced, encountered, and interpreted. According to the Pew Research Center (2025), Instagram has approximately 2 billion users worldwide and is especially popular among young adults in the United States, with 76% of individuals aged 18–24 and 57% of those aged 25–30 reporting use; platform usage is somewhat higher among women (55%) than men (44%). YouTube, with an estimated 2.5 billion users globally, reaches one of the most demographically diverse audiences, exhibiting relatively balanced representation across gender, age groups, and racial categories. Reddit, by contrast, serves a more demographically concentrated user base in the U.S., skewing younger than 40, predominantly male (29% of men vs. 23% of women in the U.S. use it), and disproportionately White or Asian.

These pronounced differences across platforms suggest that social meaning might be inherently perspectival: the same content may be perceived differently depending on platform context and the background of the individual interpreting it. Motivated by this insight, P1SCO adopts a perspectivist annotation framework that explicitly preserves annotator-level variation rather than collapsing judgments into a single "ground truth." By

[1] https://anonymous.4open.science/r/p1sco

Dimension	Keywords
Knowledge	Exchange of ideas or information; learning, teaching: teaching, intelligence, competent, expertise, know-how, insight
Power	Having power over the behavior and outcomes of another: command, control, dominance, authority, pretentious, decisions
Status	Conferring status, appreciation, gratitude, or admiration upon another: admiration, appreciation, praise, thankful, respect, honor
Trust	Will of relying on the actions or judgments of another: trustworthy, honest, reliable, dependability, loyalty, faith
Support	Giving emotional or practical aid and companionship: friendly, caring, cordial, sympathy, companionship, encouragement
Romance	Intimacy among people with a sentimental or sexual relationship: love, sexual, intimacy, partnership, affection, emotional, couple
Similarity	Shared interests, motivations or outlooks: alike, compatible, equal, congenial, affinity, agreement
Identity	Shared sense of belonging to the same community or group: community, united, identity, cohesive, integrated
Fun	Experiencing leisure, laughter, and joy: funny, humour, playful, comedy, cheer, enjoy, entertaining
Conflict	Contrast or diverging views: hatred, mistrust, tense, disappointing, betrayal, hostile

Table 1: Definitions of the 10 social dimensions by Choi et al. (2020).

combining cross-platform data with rich annotator metadata, the dataset enables systematic study of how social dimensions vary across platforms, how individual differences shape social perception, and how disagreement itself serves as a meaningful signal of social interpretation.

3 Data Collection Methodology

As a first step in the collection of our dataset, we source comments from social media. To sample diverse comments, we follow previous work by Choi et al. (2020). We randomly sample comments from three platforms: YouTube, Reddit, and Instagram.

Instagram: We source posts, comments and replies from the Instagram Influencer Dataset (Kim et al., 2020). The dataset contains posts from over 38k influencers and over 1.6M comments. We extract 20 000 comments.

YouTube: To randomly sample from YouTube, we follow the methodology proposed by Zhou et al. (2011). We use the YouTube Search API and randomly generate prefixes to search for valid IDs. We then use these IDs to randomly sample up to 100 comments from each video. We collect 20 000 comments.

Reddit: We randomly sample comments from the Pushshift Reddit dataset (Baumgartner et al., 2020). We extract 0.01% of the comments available between 2013 until 2024.

For every platform, we select first-level comments that are longer than 20 words and shorter than 100, excluding emoji. We exclude non-English comments. We sample 2000 comments from those extracted for each platform, for a total of 6000 unique comments across all platforms. While previous work selected only comments that included first- or second-person pronouns (Choi et al., 2020), to focus on direct interaction, we manually examined a sample of comments and found the filter to be unnecessary to obtain interactive comments.

3.1 Annotation Interface

We design an annotation interface using Streamlit.[2] Participants are first asked to answer a series of sociodemographic questions as well as complete the 10-question Big 5 personality questionnaire (Rammstedt and John, 2007). The participants are then guided through the definition of each dimension and some examples. To source examples, we use the original dataset and find sentences where all annotators agreed on a single dimension. Once the participants have read the instructions, they are shown the task. Participants may select any of the ten dimensions as well as 'Other', in which case they are prompted to specify. They may also select 'None'. We use form validation to ensure 'None' is not selected if other dimensions are. To test the interface, we ran several pilots, first within the authors, then with small samples of participants. Screenshots of the annotation interface are shown in the Appendix.

4 Dataset

We recruit participants from the U.K. and the U.S. using Prolific.[3] To ensure data quality while allowing diverse views, we ensure that the demographic information provided through our study matches their Prolific profile. We also remove users that look too long or too short to complete the task (5th and 95th percentile). In addition, we collect three duplicate examples, using intra-annotator agreement as a further check (Abercrombie et al., 2025). Each annotator labeled 50 unique examples. In total, after removing malicious participants, we collect 28 775 annotations by 543 participants. Our participant pool is roughly even in terms of gender and nationality. Sociodemographic distribution are

[2] https://streamlit.io
[3] https://www.prolific.com

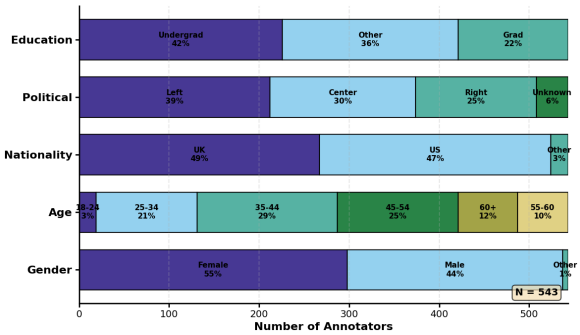


Figure 2: Demographic description of the participants.

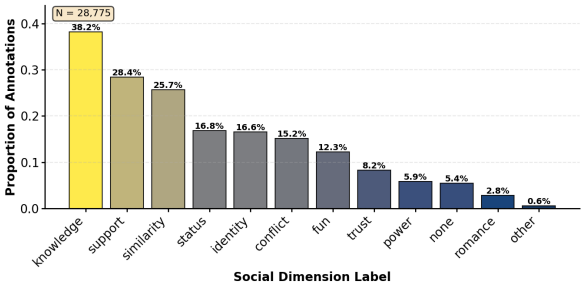


Figure 3: Overall distribution of assigned labels

shown in Figure 2. A datasheet following Gebru et al. (2021) can be found in the Appendix.

4.1 Label Analysis

Figure 3 shows the distribution of social dimension annotations across the dataset (N = 28 775). Knowledge is the most frequently assigned label (38.2%), followed by support (28.4%), and similarity (25.7%). Mid-frequency dimensions include status, identity, and conflict; while trust, power, romance, and none are comparatively rare. Overall, the distribution is highly skewed, with a small number of dimensions accounting for the majority of annotations.

As our task allows for multiple labels, we consider how commonly dimensions co-occur. Figure 4 shows the distribution of the number of dimensions present in each example. The majority of examples have at least two dimensions present: annotators frequently perceive multiple, overlapping social meanings within the same interaction.

Figure 5 shows the co-occurrence of labels and conditional probabilities across the entire dataset. Three pairs of social dimensions tend to co-occur often: identity and similarity, trust and support, and power and conflict. The first pair represents joint expressions of identity by the speakers, e.g., *I have been the lone reader until recently when I found more reader people.* The second pair indicates situations of mutual aid, sympathy, and empathy, e.g., *@user awe. So kind. Thank you. I'm hoping that sharing my mom's journey helps others.* The last pair is a typical example of a struggle, where an attempt to exert power elicits a conflictual reaction, e.g., *Y'all need to stop posting straight up copying as an off brand. It's not a substitution for the original and it's not really funny or crappy, it's just an asshole thing to do*. The figure reveals substantial asymmetry and structure in how dimensions co-occur. Several dimensions (most notably knowledge, support, similarity, and identity) frequently co-occur with a wide range of other dimensions, whereas dimensions such as romance, power, and conflict tend to co-occur more selectively. This pattern reinforces the view that social dimensions are not independent labels but interrelated interpretive constructs.

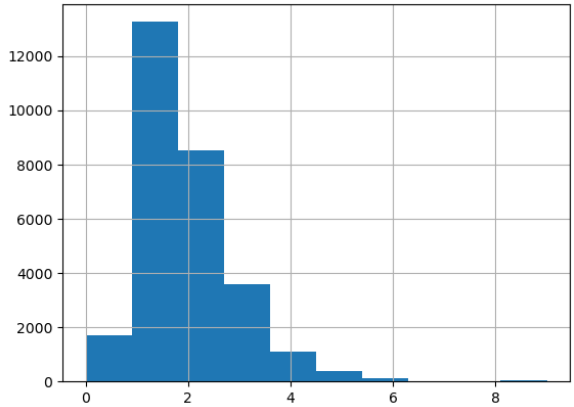
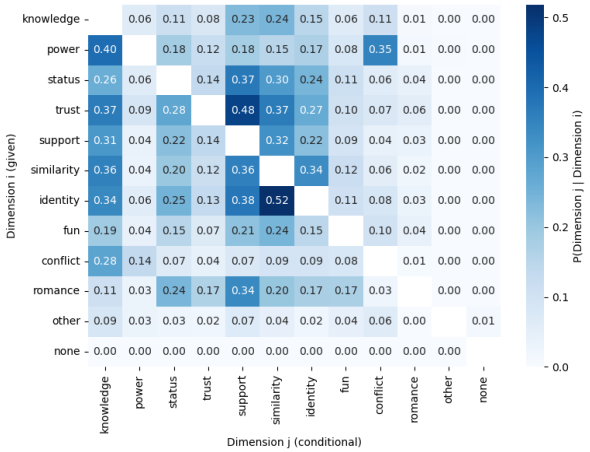


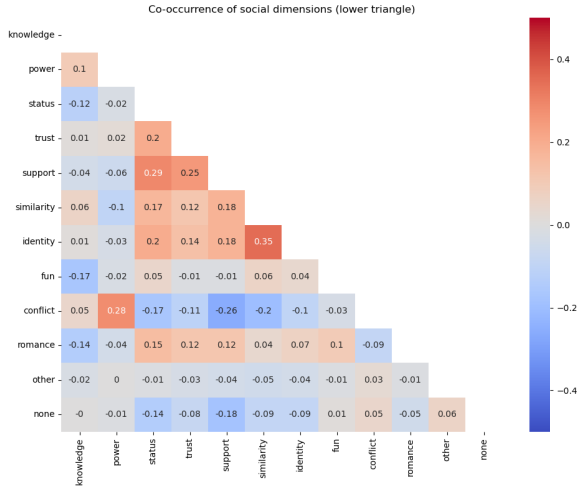
Figure 4: Distribution of number of labels per example.

Other We manually look at the examples labeled 'other'. Generally, when asked to specify, users described either the emotion they perceived, sarcasm, or they were unable to understand the dimension from the context provided.

Agreement We measure agreement using pairwise Cohen's κ, shown in Tables 2 and 5. To understand the effects of sociodemographics, we calculate agreement stratifying by groups as well as overall. We see moderate improvements in agreement, suggesting some group affinities, but agreement either remains stable or decreases in others, suggesting the task is highly subjective and likely challenging. Pairwise Cohen's κ is used over Krippendorff's α to capture localized agreement between annotators, which is more appropriate for a perspectival annotation task where systematic differences between annotators are expected and



(a) Conditional co-occurrence probabilities of social dimensions.



(b) Joint co-occurrence probabilities of social dimensions.

Figure 5: Co-occurrence matrices of social dimensions in the full dataset.

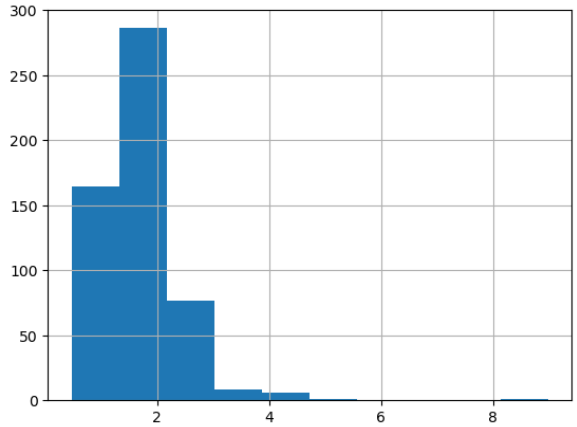


Figure 6: Distribution of average number of labels per annotator.

Label	Instagram	Reddit	YouTube	Average
other	0.992	0.976	0.971	0.980
romance	0.836	0.936	0.940	0.904
none	0.927	0.743	0.755	0.808
power	0.884	0.751	0.747	0.794
trust	0.585	0.774	0.765	0.708
fun	0.631	0.711	0.732	0.691
conflict	0.888	0.576	0.569	0.677
status	0.329	0.677	0.650	0.552
identity	0.383	0.518	0.568	0.490
support	0.190	0.545	0.511	0.415
similarity	0.214	0.391	0.454	0.353
knowledge	0.414	0.288	0.322	0.341
Average	0.606	0.657	0.665	0.643

Table 2: Pairwise Kappa by social dimension label and platform. Values represent the average pairwise kappa across all annotator pairs.

disagreement is not treated as noise. We also report Krippendorff's α in the Appendix.

Mixed Effect Model We used mixed-effects logistic regression models to examine how annotator characteristics predict label assignment. We fit separate models for each of the 10 social dimension labels (*knowledge, power, status, trust, support, similarity, identity, fun, conflict, romance;* plus *other* and *none*).

Fixed effects included demographic variables—gender (female vs. male/other), nationality (U.S. vs. U.K.), age group (reference category: most common group), education level (reference category: most common), and ethnicity (reference category: most common). We collapse both the U.K. and U.S. political spectra into left, right, or center (reference category: most common). Platform is included as a categorical predictor with three levels (Instagram, YouTube, Reddit; reference category: Instagram). In addition, we include the Big Five personality traits—Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism—as continuous predictors and standardized ($M = 0$, $SD = 1$) prior to analysis.

Models were fitted using the `mixedlm` function from the `statsmodels` package in Python, with maximum likelihood estimation (`REML = False`). Statistical significance was assessed at $\alpha = 0.05$. The coefficients are shown in Figure 7. The platform has the strongest effect, pointing to context and platform affordances and characteristics having the biggest effect on interactions. Sociodemographics show modest effects, especially gender, age, and education.

Label	YouTube					Reddit					Instagram				
	Overall	Female	Male	US	UK	Overall	Female	Male	US	UK	Overall	Female	Male	US	UK
knowledge	0.322	0.323	0.314	0.309	0.318	0.288	0.316	0.245	0.289	0.280	0.414	0.434	0.413	0.390	0.411
power	0.747	0.758	0.749	0.739	0.780	0.751	0.745	0.755	0.752	0.756	0.884	0.883	0.876	0.868	0.896
status	0.650	0.674	0.582	0.619	0.674	0.677	0.684	0.645	0.615	**0.741**	0.329	0.361	0.298	0.311	0.368
trust	0.765	0.779	0.754	0.748	0.773	0.774	0.793	0.740	0.765	0.765	0.585	0.614	0.562	0.569	0.609
support	0.511	0.524	0.474	0.468	0.548	0.545	0.586	0.502	0.505	0.566	0.190	0.206	0.170	0.175	0.203
similarity	0.454	0.491	0.388	0.471	0.446	0.391	0.417	0.340	0.416	0.379	0.214	0.234	0.200	0.237	0.188
identity	0.568	0.603	0.510	0.555	0.584	0.518	0.552	0.466	0.545	0.500	0.383	0.403	0.361	**0.437**	0.356
fun	0.732	0.734	0.718	0.711	0.757	0.711	0.710	0.719	0.706	0.730	0.631	0.639	0.642	0.595	0.654
conflict	0.569	0.569	0.559	0.537	0.611	0.576	0.561	0.605	0.546	0.606	0.888	0.900	0.880	0.880	0.885
other	0.971	0.978	0.966	0.981	0.962	0.976	0.984	0.960	0.975	0.979	0.992	0.993	0.990	0.992	0.993
none	0.755	0.743	0.773	0.781	0.724	0.743	0.718	0.784	0.758	0.728	0.927	0.915	0.940	0.923	0.928
romance	0.940	0.949	0.929	0.926	0.963	0.936	0.937	0.934	0.914	0.954	0.836	0.854	0.802	0.830	0.839

Table 3: Within-group agreement with pairwise Cohen's κ: gender and nationality by platform. Bold values indicate κ substantially higher (> 0.05) than overall agreement for that platform.

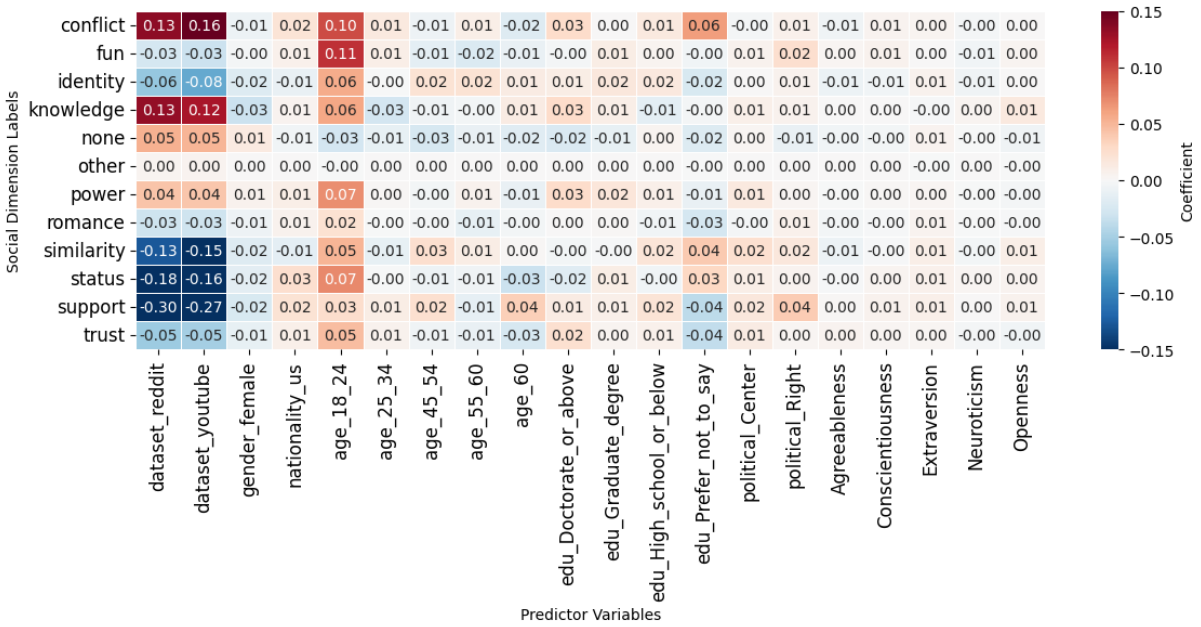


Figure 7: Coefficients for the mixed effects model. For the source platform, we use Instagram as a reference. For the other groups, we use the subgroup with the largest sample.

4.2 Platform Differences

Given the differences in affordances and audience, we expect to see variation in the distribution of labels between platforms. Figure 8 shows the label distributions for each platform. YouTube and Reddit show the most similar distributions: knowledge is the most dominant dimension present, followed by similarity, support, and conflict. In contrast, Instagram interactions are dominated by support, similarity, knowledge, and status. Romance remains low across platforms, but is significantly more common on Instagram. Instagram also shows a significantly smaller proportion of posts with no dimensions present. In contrast to Instagram, where support and similarity dominate, Reddit and YouTube are more strongly characterized by knowledge and conflict, pointing to different interactional expectations across platforms. These differences reflect not only variation in platform interaction styles, but also how annotators interpret social meaning in context, with platform cues influencing which dimensions are perceived as salient.

5 Sociodemographic Differences

Given the findings of the previous model, we explore more in depth the sociodemographic differences in labels. We find that annotators under 25, males, and US-nationals tend to assign a larger number of labels per example than other demographics groups. In addition, gender and personality traits show effects on label distribution:

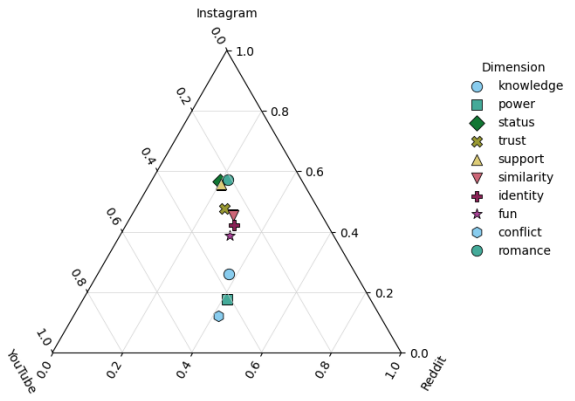


Figure 8: Proportion of each dimension per platform.

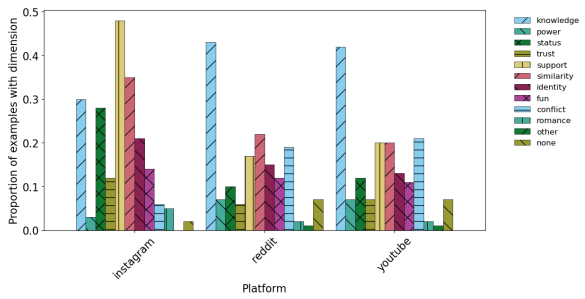


Figure 9: Distribution of presence of each dimension for each platform.

5.1 Gender

Debora Tannen presented *genderlects*, the idea that men and women use language differently, with women using language to build rapport and men to show and gain power (Tannen, 1990). While Tannen's original claims concern language production, we extend this logic to perception: if gendered socialisation shapes how one uses language, it plausibly shapes how one interprets it. We therefore ask whether male and female annotators systematically differ in which social dimensions they perceive, a perceptual analogue of the genderlect hypothesis.

We first check whether men and women assign the same number of labels overall. On average, men assign more labels per example than women (see Table 4). This difference is statistically significant (Mann-Whitney-U test, $p < .001$).

Gender	Mean	Median	Std. Dev.	Count
Female	1.698	1.0	0.949	15 788
Male	1.828	2.0	1.101	12 669

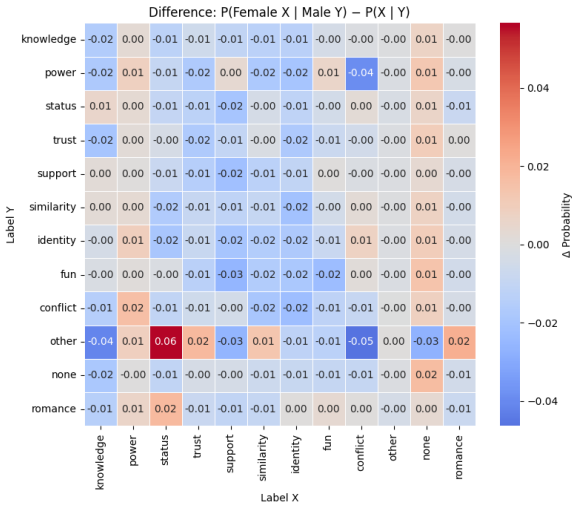
Table 4: Descriptive statistics for the number of labels assigned by the annotators disaggregated by gender.

A chi-square test comparing the distribution of applied labels by gender is significant, $\chi^2(10) = 143.33$, $p < .001$, indicating that men and women differed not only in the number of labels assigned but also in the relative frequency with which specific labels are applied. This analysis conditions on the total number of labels assigned, isolating differences in label selection from overall labeling propensity. When analyzing each platform, gender differences in label selection were larger on Reddit ($p < .001$) and YouTube ($p < .001$), while the effect on Instagram was weaker ($p = .033$). These results suggest that gendered annotation strategies interact with platform-specific content characteristics rather than reflecting a uniform difference across platforms.

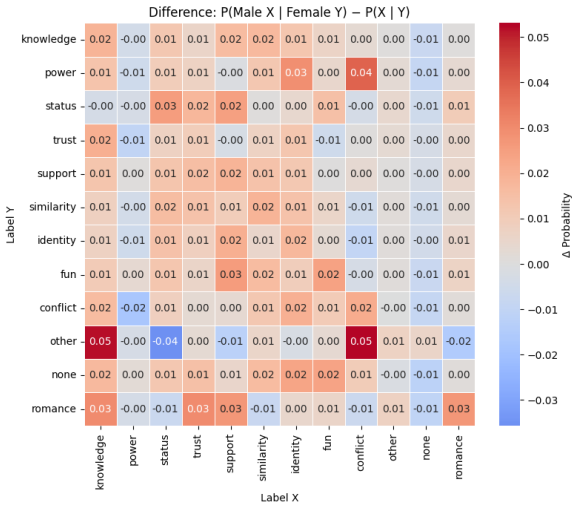
To understand how the labels differ, we calculate conditional co-occurrence probabilities (see Figure 10). The diagonal of the matrices shows overall agreement between men and women. Notably, women often label an item as other when men assign status. In turn, women are less likely to label something as power when men have labeled it as conflict or other. Conversely, men are more likely to label other women's conflict as other or power, and to label romance where women see knowledge, trust, or support. This pattern is directionally consistent with Tannen's characterization of male-coded and female-coded communicative orientations, now observed at the level of social perception rather than production. However, the effect sizes are modest and interact with platform, suggesting that gendered interpretive schemas are real but contextually modulated.

5.2 Big 5 Personality Traits

Figure 11 shows the Pearson correlations between annotators' Big Five personality traits and the proportion of social dimensions they assigned. Overall, correlations are small in magnitude, indicating no strong linear relationship between personality traits and annotation behavior. A small number of weak but statistically significant associations are observed. For example, openness shows a modest positive correlation with knowledge annotations, while conscientiousness is weakly associated with support. Conversely, the 'other' category is negatively correlated with both openness and conscientiousness. These patterns suggest that annotator characteristics may subtly influence which social dimensions are foregrounded, without implying systematic bias or determinism in annotation deci-



(a) Conditional probabilities: how likely is a woman to label an example X given that a man labelled it Y



(b) Conditional probabilities: how likely is a man to label an example X given that a woman labelled it Y

Figure 10: Conditional probabilities between genders.

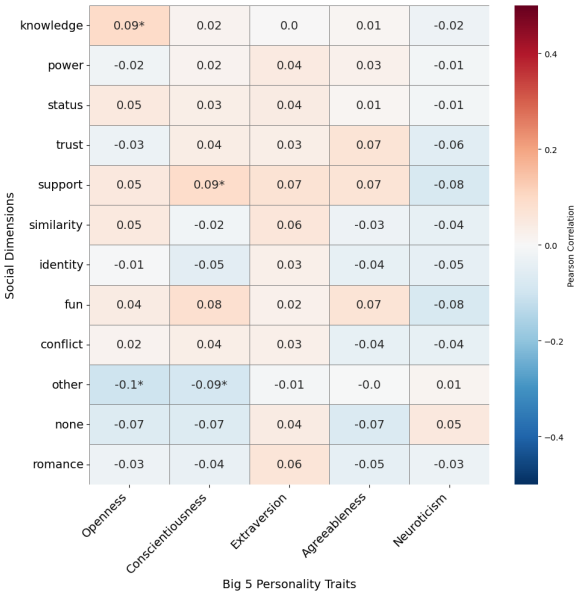


Figure 11: Correlations between Big 5 personality traits and label assignments for each dimension. Significance is Bonferroni-corrected.

sions. Given the small effect sizes and correlational nature of the analysis, these results should be interpreted as exploratory rather than explanatory.

6 Discussion and Conclusion

This work provides empirical evidence for perspectivism in social dimension annotation, demonstrating that demographic characteristics, personality traits, and platform context systematically shape how annotators perceive and label social interactions. Through mixed-effects modeling of over 28k annotations, we show that individual differences are a meaningful signal reflecting genuine diversity in social perception.

Our key contributions are: we collect the first dataset of perspectivist unaggregated annotations of social dimensions; through this dataset, we demonstrate that demographic factors (gender, age, political ideology, nationality) and Big Five personality traits significantly predict label assignment across the 10 social dimensions, with effects consistent across Instagram, YouTube, and Reddit. We show that homogeneous demographic groups exhibit higher within-group agreement than the overall population, supporting the premise that shared experiences create convergent interpretive frameworks; we provide methodological guidance for perspectivist annotation, including preserving individual perspectives, modeling annotator characteristics explicitly, and reporting demographic composition transparently.

Our findings contribute to a growing body of work acknowledging human factors in computational humanities, and our dataset enables future work to develop new methods and models for human-centered studies. Future work should extend this framework to additional languages and cultures, explore intersectional demographic effects, and develop computational models that explicitly represent multiple viewpoints. The dataset and code are available at `https://anonymous.4open.science/r/p1sco` to facilitate such research.

Limitations

Annotation quality Social dimensions are inherently interpretive, and annotations reflect annotators' perceptions rather than objective properties of text or the dimensions intended by the commenters. While this perspectival approach is a central contribution of the dataset, it also limits the extent to which labels can be treated as ground truth or used for definitive classification tasks. Moreover, annotating ten possible dimensions is a cognitively difficult task. While annotators assigned social dimension labels, they did not provide explicit rationales for their choices. This limits insight into the reasoning behind disagreements and constrains fine-grained analysis of interpretive variation.

Diversity While we took care to expand our pool of annotators, our pool was not ethnically diverse and was focused on the United States and the United Kingdom. As a result, the dataset reflects a bounded set of perspectives, and social meanings may be interpreted differently by annotators from other cultural or socio-linguistic contexts. Future work could expand this pool to be more diverse and to cover languages other than English.

Generalizability The dataset consists of content from a small number of social media platforms, but our study shows platform has an important impact on the labels. This may limit our findings generalizability to other social media platforms.

Label Distribution Several social dimensions are relatively rare, leading to skewed label distributions. This affects the reliability of agreement metrics and limits the usefulness of the dataset for training balanced supervised models without additional resampling or aggregation.

Ethical Considerations

Annotators were paid the hourly wage recommended by the platform (9 GBP/hour). While we took care to filter upsetting material (in particular, pornographic bots on Reddit), annotators were warned that the material they may see could be offensive in nature given the sources of data. The dataset is available fully anonymized.

Acknowledgments

AI-based assistants were used to provide coding, writing and editing assistance during manuscript preparation. All content, analyses, and conclusions were reviewed and verified by the authors.

References

Gavin Abercrombie, Tanvi Dinkar, Amanda Cercas Curry, Verena Rieser, and Dirk Hovy. 2025. Consistency is key: Disentangling label variation in natural language processing with intra-annotator agreement. In *Proceedings of the The 4th Workshop on Perspectivist Approaches to NLP*, pages 63–74, Suzhou, China. Association for Computational Linguistics.

Luca Maria Aiello, Daniele Quercia, Ke Zhou, Marios Constantinides, Sanja Šćepanović, and Sagar Joglekar. 2021. How epidemic psychology works on twitter: Evolution of responses to the covid-19 pandemic in the us. *Humanities and social sciences communications*, 8(1).

Duilio Balsamo, Paolo Bajardi, Gianmarco De Francisci Morales, Corrado Monti, and Rossano Schifanella. 2023. The pursuit of peer support for opioid use recovery on reddit. In *Proceedings of the international AAAI conference on web and social media*, volume 17, pages 12–23.

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.

Simon Martin Breum, Daniel Vædele Egdal, Victor Gram Mortensen, Anders Giovanni Møller, and Luca Maria Aiello. 2024. The persuasive power of large language models. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, pages 152–163.

Federico Cabitza, Andrea Campagner, and Valerio Basile. 2023. Toward a perspectivist turn in ground truthing for predictive computing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 6860–6868.

Minje Choi, Luca Maria Aiello, Krisztián Zsolt Varga, and Daniele Quercia. 2020. Ten social dimensions of conversations and relationships. In *Proceedings of The Web Conference 2020*, pages 1514–1525.

Sebastian Deri, Jeremie Rappaz, Luca Maria Aiello, and Daniele Quercia. 2018. Coloring in the links: Capturing social ties as they are perceived. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–18.

Simona Frenda, Gavin Abercrombie, Valerio Basile, Alessandro Pedrani, Raffaella Panizzon, Alessandra Teresa Cignarella, Cristina Marco, and Davide Bernardi. 2025. Perspectivist approaches to natural language processing: a survey. *Language Resources and Evaluation*, 59(2):1719–1746.

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92.

Dirk Hovy and Diyi Yang. 2021. The importance of modeling social factors of language: Theory and practice. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies*, pages 588–602.

Seungbae Kim, Jyun-Yu Jiang, Masaki Nakada, Jinyoung Han, and Wei Wang. 2020. Multimodal post attentive profiling for influencer marketing. In *Proceedings of The Web Conference 2020*, pages 2878–2884.

Lucio La Cava, Luca Maria Aiello, and Andrea Tagarelli. 2023. Drivers of social influence in the twitter migration to mastodon. *Scientific Reports*, 13(1):21626.

Lorenzo Lucchini, Luca Maria Aiello, Laura Alessandretti, Gianmarco De Francisci Morales, Michele Starnini, and Andrea Baronchelli. 2022. From reddit to wall street: The role of committed minorities in financial collective action. *Royal Society Open Science*, 9(4):211488.

Corrado Monti, Luca Maria Aiello, Gianmarco De Francisci Morales, and Francesco Bonchi. 2022. The language of opinion change on social media under the lens of communicative action. *Scientific Reports*, 12(1):17920.

Matthias Orlikowski, Jiaxin Pei, Paul Röttger, Philipp Cimiano, David Jurgens, and Dirk Hovy. 2025. Beyond demographics: Fine-tuning large language models to predict individuals' subjective text perceptions. *CoRR*.

Cecilia Ovesdotter Alm. 2011. Subjective natural language problems: Motivations, applications, characterizations, and implications. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 107–112, Portland, Oregon, USA. Association for Computational Linguistics.

Jiaxin Pei and David Jurgens. 2023. When do annotator demographics matter? measuring the influence of annotator demographics with the popquorn dataset. In *Proceedings of the 17th linguistic annotation workshop (LAW-XVII)*, pages 252–265.

Pew Research Center. 2025. Social media fact sheet. https://www.pewresearch.org/internet/fact-sheet/social-media/#who-uses-each-social-media-platform. Accessed: 2026-01-13.

Barbara Plank. 2022. The "problem" of human label variation: On ground truth in data, modeling and evaluation. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 10671–10682, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Beatrice Rammstedt and Oliver P. John. 2007. Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of Research in Personality*, 41(1):203–212.

Deborah Tannen. 1990. Gender differences in conversational coherence: Physical alignment and topical cohesion. In *This chapter is a slightly revised and shortened version of a paper presented at the" Gender Differences in Conversational Interaction" panel at the 1988 Georgetown University Round Table on Languages and Linguistics, Washington, DC, Mar 1988*. Ablex Publishing.

Jia Zhou, Yanhua Li, Vijay Kumar Adhikari, and Zhi-Li Zhang. 2011. Counting youtube videos via random prefix sampling. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 371–380.

A Datasheet for the P1SCO Dataset

Motivation

Language is inherently ambiguous with different groups and individuals bring their own flavor to it. This dataset aims to enable researchers to study diversity in the expression and perception of pragmatics.

Composition

The dataset contains 28k annotated text-only comments from Instagram, YouTube and Reddit. The dataset is labeled with the ten social dimensions, other or none, as well as sociodemographic profiles and personality of the annotators. We do not know the demographic profiles of the creators of the data that was annotated. See main body of paper for more detailed demographics of the annotators.

Collection Process

Described in the paper. Crowdworkers are paid the wage recommended by Prolific of 9 GBP/hour. We based the time estimate on in-house pilots completed by the researchers. We adjusted the payment to crowdworkers to match the recommended payment.

Annotation Process

Described in the paper. Annotation interface shown in Appendix C.

Uses

The intended use of this dataset is to study and model diversity of perspectives in pragmatics, both at the individual and group level. The dataset will also enable the development of new methods for perspectivist models and frameworks. The dataset may also be used to develop models that enable researchers to study social interactions over social media at a more fine-grained detail and with diverse perspectives and to benchmark models against diverse perspectives.

Distribution

The dataset is available at https://anonymous.4open.science/r/p1sco/README.md under an MIT license. The dataset should not be used to profile or otherwise violate the rights to dignity and autonomy of individuals or groups.

Maintenance

Anonymized for submission.

Ethical Considerations and Limitations

The main limitations are discussed in the paper. In terms of ethical considerations, the dataset has been anonymized by removing usernames and reassigning worker IDs to crowdworkers. The dataset is mainly annotated by U.S. and U.K. citizens, who are mostly white. In addition, the dataset labels show the perspective of a listener, not of the speaker.

B Agreement

Table 5: Within-group agreement: gender and nationality by platform. Bold values indicate α substantially higher (> 0.05) than overall agreement for that dataset.

Label	YouTube					Reddit					Instagram				
	Overall	Female	Male	US	UK	Overall	Female	Male	US	UK	Overall	Female	Male	US	UK
knowledge	0.30	0.31	0.29	0.31	0.29	0.27	0.28	0.24	0.26	0.27	0.30	0.31	0.31	0.27	0.31
power	0.08	0.10	0.03	0.11	0.08	0.08	**0.13**	0.04	0.09	0.09	0.04	0.04	0.02	0.04	0.04
status	0.17	0.17	0.13	0.15	0.18	0.12	0.14	0.08	0.06	0.16	0.16	0.19	0.13	0.15	0.18
trust	0.05	0.06	0.07	0.05	0.03	0.06	0.05	0.07	0.02	0.09	0.02	0.03	0.00	0.00	0.04
support	0.26	0.26	0.24	0.23	0.26	0.22	0.25	0.20	0.19	0.26	0.20	0.21	0.17	0.18	0.21
similarity	0.15	0.17	0.12	0.16	0.14	0.12	0.13	0.11	0.10	0.13	0.14	0.15	0.15	0.15	0.12
identity	0.09	0.08	0.12	0.08	0.13	0.09	0.08	0.08	0.07	0.09	0.07	0.09	0.03	0.06	0.07
fun	0.30	0.29	0.29	0.31	0.30	0.29	0.28	0.29	0.27	0.32	0.26	0.27	0.24	0.23	0.27
conflict	0.36	0.36	0.34	0.36	0.38	0.32	0.27	0.36	0.32	0.32	0.44	0.47	0.43	0.41	0.43
romance	0.13	0.05	0.16	0.12	0.16	0.18	0.21	0.20	0.09	**0.26**	0.12	0.12	0.08	0.15	0.14
other	0.02	-0.01	-0.01	-0.00	0.04	-0.01	-0.00	-0.01	-0.00	-0.01	-0.00	-0.00	-0.00	-0.00	-0.00
none	0.06	0.07	0.04	0.05	0.07	0.04	0.05	0.03	0.02	0.05	0.03	0.05	0.01	0.03	**0.10**

C Data collection interface

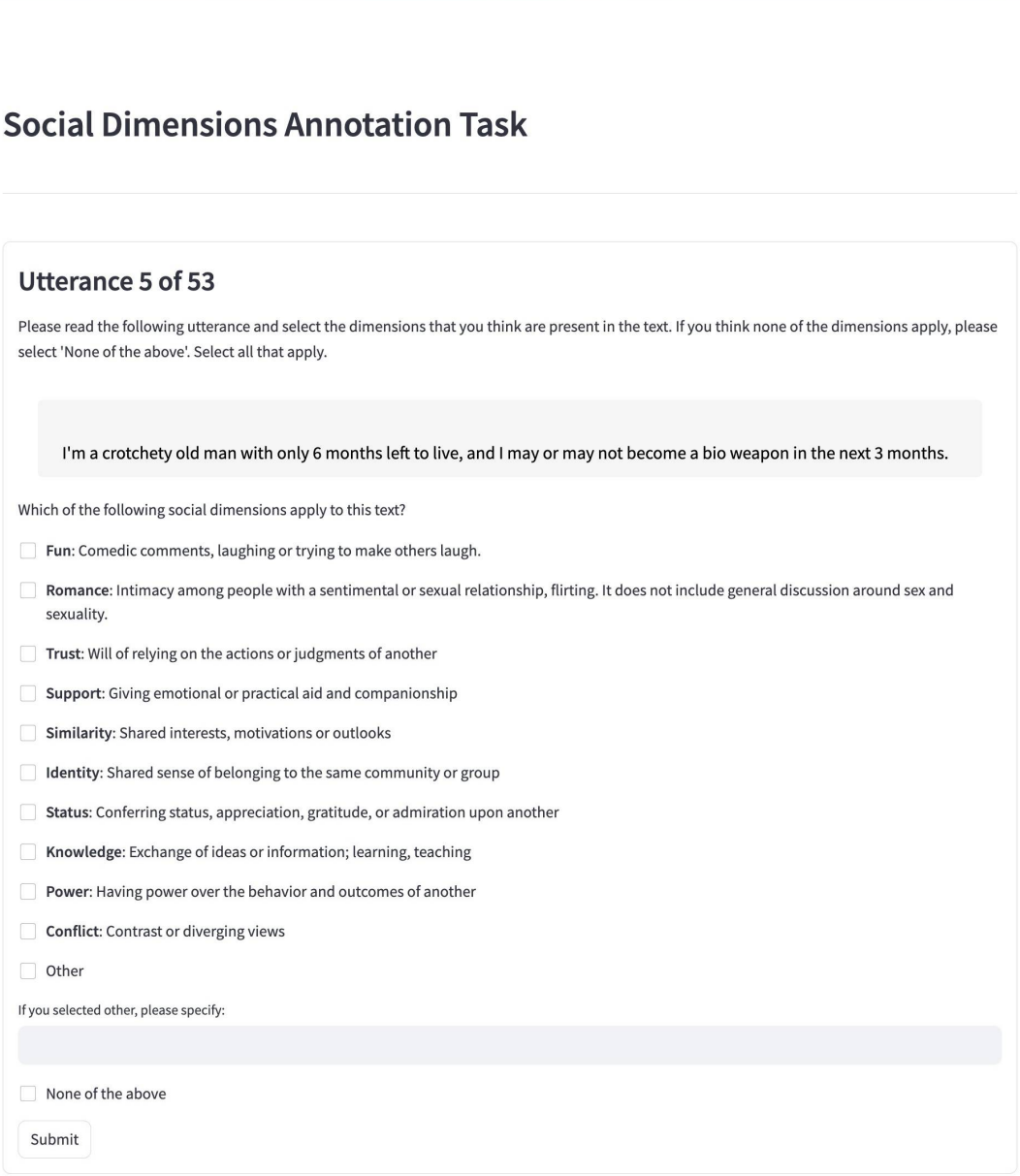


Figure 12: Screenshot of the P1SCO annotation interface. Participants were shown one utterance at a time and asked to select all applicable social dimensions. A free-text “Other” field and a “None of the above” option were also provided.