

Detecting Perspective-Getting in Wikipedia Discussions

Evgeny Vasilets[♡] Tijs Van den Broek[♣] Anna Wegmann[♠]

David Abadi[♡] Dong Nguyen[♠]

[♡] University of Amsterdam, The Netherlands

[♣] VU Amsterdam, The Netherlands

[♠] Utrecht University, The Netherlands

t.a.vanden.broek@vu.nl, a.m.wegmann@uu.nl

Abstract

Perspective-getting (i.e., the effort to obtain information about the other person’s perspective) can lead to more accurate interpersonal understanding. In this paper, we develop an approach to measure perspective-getting and apply it to English Wikipedia discussions. First, we develop a codebook based on perspective-getting theory to operationalize perspective-getting into two categories: asking questions about and attending the other’s perspective. Second, we use the codebook to annotate perspective-getting in Wikipedia discussion pages. Third, we fine-tune a RoBERTa model that achieves an average F-1 score of 0.76 on the two perspective-getting categories. Last, we test whether perspective-getting is associated with discussion outcomes. Perspective-getting was not higher in non-escalated discussions. However, discussions starting with a post attending the other’s perspective are followed by responses that are more likely to also attend the other’s perspective. Future research may use our model to study the influence of perspective-getting on the dynamics and outcomes of online discussions.

1 Introduction

Understanding what other individuals think or feel is important for effective interpersonal behavior, such as building and maintaining social relationships (Hughes and Leekam, 2004; Keysar et al., 2003), effective customer service (Axtell et al., 2007), or reaching agreements in negotiations (Galinsky et al., 2008; Trötschel et al., 2011). Individuals take another individual’s perspective (i.e., engage in *perspective-taking*) by putting themselves in the same “mental shoes” as the other (Galinsky et al., 2008) and by recognizing that this other person’s perspective differs from their ego-centric perspective (Epley et al., 2004a). Scholars have shown that imagining someone else’s

perspective and incorporating this alternative viewpoint in communication is a notoriously difficult task (Eyal et al., 2018; Epley et al., 2004a; Damen et al., 2020, 2019). Individuals may take their own perspective as an anchor for assessing other’s perspective, lack motivations to seek for information about other’s perspective, or consider perspective-taking as requiring too much mental effort (Epley et al., 2004a,b). Consequently, individuals often fail to accurately infer the perspective of others (Realo et al., 2003; Eyal et al., 2018).

In contrast to the cognitive process of imagining an other’s perspective through perspective-taking, the strategy of **perspective-getting** results in a more accurate assessment of the other’s perspective (Bruneau and Saxe, 2012; Eyal et al., 2018; Kalla and Broockman, 2023). Perspective-getting refers to directly asking about and listening to information related to another person’s thoughts, feelings, beliefs, or preferences (Damen et al., 2021; Eyal et al., 2018; Kalla and Broockman, 2023) instead of simply imagining the other’s perspective. In contrast to perspective-taking, perspective-getting can likely be more accurately annotated and detected as it is about detecting visible behavior as opposed to cognitive states from text. Perspective-getting has been operationalized as whether the interlocutor asks a target interlocutor about their perspective (Eyal et al., 2018; Damen et al., 2021) and as whether interlocutors summarize the other’s perspective (Kalla and Broockman, 2023).

Previous work manipulated perspective-getting via participant instructions in experimental settings (Bruneau and Saxe, 2012; Eyal et al., 2018; Damen et al., 2021; Kalla and Broockman, 2023). To the best of our knowledge, perspective-getting was not studied outside lab settings and has not been investigated computationally. We study perspective-getting in a natural online discussion settings on Wikipedia and build a classifier for computational analysis.

	Shortened Example
Attending the Other’s Perspective (AOP)	<p>✓ Message: MOS:INTRO states for the LEDE explicit "Do not hint at startling facts without describing them". I suppose that the "War beginning" is such a startling facts , @DS: why do you refuse to give the reader the information about issues 1-4 in the lede?.</p> <p>Reply: To state when the war started is not a "startling fact". I see no reason to mention in the lede that the Chilean troops capturing Antofagasta got a warm welcome, it is quite simply an undue bit of trivia for the lede.</p>
	<p>✗ Message: If the statement is not in serious dispute (i.e. in high-quality sources) then it’s a fact we can just assert. Failing to do so would not be neutral. </p> <p>Reply: We’ve gotten circular at this point, need input from others.</p>
Asking for Clarification (AFC)	<p>✓ Message: This article barely reads like an article. Opinions are stated as facts, speculations are presented as conclusions, etc., and numerous questions are omitted.</p> <p>Reply: It would help if you were more specific.</p>
	<p>✗ Message: Isabel Allende is the Spanish speaking equivalent to English author E.L. James. She is the author of the fifty shades of grey books. I’m curious, would you use a biography of E.L. James to validate an assertion on a war history article? I’m going out on a limb here but it appears your clearly out of you depth when comes to Latin american history articles, the fact your clearly using a quote finder and not being scrupulous with the quote and the source validates my point.</p> <p>Reply: You do not get to dismiss a source just because you do not like it, not a single piece of text you have written actually disputes the sources I have supplied, cheers.</p>

Table 1: *Perspective-getting*. For our perspective-getting categories AOP and AFC, we display two example messages, one displaying (✓) and one not displaying (✗) the category. We also display the previous message (Message) that the considered message replies to (Reply). Annotators see both messages during annotation, RoBERTa only sees the Reply. Perspective-getting refers to AFC about the other’s perspective (Damen et al., 2021; Eyal et al., 2018) and listening to the other’s perspective — which was operationalized with summarization by Kalla and Broockman (2023) and Bruneau and Saxe (2012). Our AOP category encompasses summarization.

Approach and contributions In this study, we aim to measure perspective-getting in messages in online discussions.

First, we develop a codebook to operationalize perspective-getting into two categories: Attending an Other’s Perspective (AOP) and Asking For Clarification (AFC). Our codebook is grounded in previous literature on perspective-getting (Kalla and Broockman, 2023; Eyal et al., 2018; Damen et al., 2021) and contains examples from Wikipedia discussions. We specifically focus on discussions between editors on Wikipedia, during which editors have to reach a consensus on a difference in point of view (Black et al., 2011; De Kock and Vlachos, 2021, 2022). We expect this to be a relevant setting to study perspective-getting as, to come to a joint resolution, it should be important to understand each other’s perspective.

Second, we annotate perspective-getting in Wikipedia discussions and then fine-tune a RoBERTa model (Liu et al., 2019), which achieves a macro average F1 score of 0.76.

Third, we test whether perspective-getting scores are associated with (i) the escalation of Wikipedia discussions to a moderator and (ii) the development of perspective-getting in Wikipedia discussions. Perspective-getting was not higher in non-

escalated discussions. However, discussions starting with a post attending the other’s perspective (AOP) are more often followed by responses that show AOP as well. The codebook, dataset and code are included in our project’s [OSF repository](#).

2 Background

Perspective-getting (Eyal et al., 2018; Damen et al., 2021; Kalla and Broockman, 2023) has been a more recent term to differentiate the cognitive process of perspective-taking from strategies that aim at “getting” the perspective from another person (e.g., directly asking about their perspective). This understanding of perspective-getting has previously been and sometimes still is incorporated in some definitions of perspective-taking (Bruneau and Saxe, 2012; Kalla and Broockman, 2020, 2023). In this work, we focus on perspective-getting but we include insights from “perspective-taking” studies if they operationalize it similar to our understanding of perspective-getting as defined in Section 3.

Perspective-getting has not yet been measured using NLP methods. But our two categories for perspective-getting (AOP and AFC, see also Table 1 and Section 3) are related to various research topics in NLP. Attending the Other’s Perspective (AOP) can manifest in various ways, including

quoting, paraphrasing or summarizing the other’s message. In NLP, various studies have investigated quotation detection (Pareti et al., 2013), paraphrase detection (Dolan and Brockett, 2005; He et al., 2020; Kanerva et al., 2023), and summarization (El-Kassas et al., 2021; Tugener et al., 2021). However, AOP can not be fully represented by any of them and even a combination would face new challenges in the dialog setting.

Asking for Clarification (AFC) is related to detecting questions, which can be challenging in online messages as, for example, question marks can be absent or questions might be phrased as information requests (Wang and Chua, 2010). Transformer-based architectures have been used with moderate success to detect questions in informal online texts (Saha et al., 2020, 2021; Kalouli et al., 2021). For clarification questions specifically, there exist corpora (Rao and Daumé III, 2018; Xu et al., 2019) and generative models (Zamani et al., 2020; Rao and Daumé III, 2019). However, our problem is even more complex than detecting clarification questions. We are interested in questions aimed at understanding the other’s perspective. Some “clarification questions” might not aim to do that (e.g., “You’re joking, right?” in Table 5).

Our categories have also been relevant in other work: For example, AFC and paraphrasing/summarizing operationalized in the AOP category have been found to be critical to establish a common ground of shared knowledge (Clark and Brennan, 1991; Clark, 1996; Krauss and Fussell, 1991) and increasing understanding of each other (Rogers and Farson, 1957).

A related task is empathy detection (Khanpour et al., 2017; Xiao et al., 2012; Gibson et al., 2015; Zhou et al., 2021; Buechel et al., 2018; Barriere et al., 2022; Lahnala et al., 2022). For example, Gibson et al. (2015) predicted the empathy level of therapists in motivational interviews and defined empathy as: “the extent to which the therapist understands and/or makes an effort to grasp the client’s perspective”. Furthermore, Sharma et al. (2020) developed a framework to measure expressed empathy in online mental health dialogs. Two categories, Explorations (acquiring more information about others’ internal states) and Interpretations (expressing understanding of states of others), relate to perspective-getting but are still too far removed from the setting of online discussions.

3 Operationalizing Text-Based Perspective-Getting

Perspective-getting refers to directly asking about (Eyal et al., 2018; Damen et al., 2021) or listening to (Eyal et al., 2018; Kalla and Broockman, 2023) other person’s perspective (Eyal et al., 2018; Kalla and Broockman, 2023), for example, their experiences, thoughts, feelings, beliefs, or preferences. In lab experiments, perspective-getting has typically been operationalized (1) as whether the perspective-getter asked about the perspective of the other (Eyal et al., 2018; Damen et al., 2021) and (2) as whether the perspective-getter summarized the other’s perspective (Bruneau and Saxe, 2012; Kalla and Broockman, 2023).

Based on (1) and (2), we introduce a codebook (see Appendix A) for the text-based detection of perspective-getting with two distinct categories: Asking for Clarification of the other’s perspective (AFC) or Attending to the Other’s Perspective (AOP), for example, through summarization. Examples of messages displaying each of the two perspective-getting categories can be found in Table 1. While each category could be more or less pronounced in a text, we represent them as a binary value in this work: present or absent. To the best of our knowledge, we are the first to operationalize text-based perspective-getting. Although our categories are intended to be generally applicable, the added examples in the annotation codebook are specific to the Wikipedia discussion pages. We now discuss our categories in more detail.

Attending the Other’s Perspective (AOP) This category is about whether the text message attends the other’s perspective (AOP). The manifestation of AOP can range from quoting or summarizing the original message (Kalla and Broockman, 2023), to otherwise mentioning or referencing specific parts of the message when engaging with the other’s perspective (e.g., by taking a stance on it). Perspective-getting and -taking literature has shown that explicitly attending the other’s perspective (e.g., summarizing or paraphrasing the other) increases the accuracy of understanding the other’s perspective (Bruneau and Saxe, 2012; Kalla and Broockman, 2023). Note that AOP includes less prevalent (e.g., simply referencing and acknowledging the other’s perspective) and more prevalent forms (e.g., actively engaging with the other perspective). In this work, both forms are part of this category; we leave more fine-grained distinctions to future work.

Type		initial	no messages 1000+ words	no errors	only complete discussions	ours
✓	# messages	4,441	4,181	3,693	1,339	263
	# discussions	217	208	203	96	20
✗	# messages	133k	123,712	98,491	14,419	1,411
	# discussions	9,006	8,768	8,711	2,834	263

Table 2: *Data Pre-Processing Steps*. We display the number of discussions and messages that remain for the escalated (✓) and non-escalated discussions (✗), after the following pre-processing steps: (1) removing messages with more than 1000 words, (2) removing duplicates or dangling messages (no errors) and (3) removing incomplete discussions. Out of these, we annotated a selection of messages (ours) that showed disagreement.

Asking for Clarification (AFC) The second category is about actively “getting” or understanding the perspective of the other. This is mostly represented through asking questions (as operationalized in previous perspective-getting literature (Damen et al., 2021; Eyal et al., 2018)). We allow the “questions” to be implicit (e.g., “I think you meant ...” instead of “Did you mean ...?”). Asking about the other’s perspective can lead to better understanding of the other’s perspective (Damen et al., 2021).

4 Dataset

WikiDisputes We use our codebook to annotate perspective-getting in the *WikiDisputes* dataset from De Kock and Vlachos (2021). This dataset is derived from Wikipedia talk pages, where editors discuss possible changes to Wikipedia articles. WikiDisputes consists of 133,019 messages in *non-escalated* discussions and 4,441 messages in *escalated* discussions, i.e. where participants asked for the discussion to be mediated by a moderator. When editors do not reach consensus, the conflict is escalated and resolved by arbitration.¹ Previous research shows that when task-related conflicts escalate, it decreases the group performance related to editing the Wikipedia page (Arazy et al., 2013).

Data filtering We performed several steps to filter the dataset. See Table 2 for the number of remaining messages and discussions after each step.² First, we removed messages with more than 1000 words to reduce annotation time (the mean number of words in a message in the full dataset was 86.19). Second, we removed message duplicates and messages that replied to a non-existent initial message (i.e., the id of the ‘parent’ message points to a message that is not in the dataset). Third, because of

our interest in discussion outcomes we removed incomplete discussions, by removing all discussions in which at least one message was removed in the first two steps. This step led to a large reduction in the number of discussions (Table 2).

Finally, a message was selected for annotation if (i) the message was a reply to a discussion partner’s message (so we did not evaluate topic-starters’ initial messages) and (ii) if the message demonstrated a form of disagreement. When there is agreement between the messages, we assume no perspective-getting to be necessary as the ‘perspective’ is already shared.³ If a person wrote more than one message in a row and directed them to the same user, these messages were merged together.

Annotating perspective-getting Based on our codebook, two researchers independently annotated the WikiDisputes messages. Both annotators were male, one being a master’s student in computational social science and the other an assistant professor in social science. The annotators each independently labeled 165 messages across three iterations. They labeled the perspective-getting of a message with respect to the previous replied-to message. After each iteration the messages that were rated differently were discussed. While the conceptual definitions of the categories remained the same, we used the insights from these discussions to adjust the illustrative examples in the codebook. The 165 messages were not used for the subsequent data analysis. In the final iteration, the Cohen’s kappas for the separate categories were: AOP = 0.78 and AFC = 0.71. Our values indicate substantial to excellent annotator agreement (Landis and Koch, 1977).

After the first three iterations, the two annotators labelled the messages separately. In our final

¹https://en.wikipedia.org/wiki/Wikipedia:Dispute_resolution

²The initial #discussions differs from De Kock and Vlachos (2021) because they used different inclusion criteria.

³Perspective-getting can also take place in discussions when participants already agree. However, such cases were not the focus of this study.

	\sum	AOP			AFC		
	F1	P	R	F1	P	R	F1
LR	0.68	0.95	0.73	0.83	0.54	0.51	0.52
RB	0.76	0.78	0.71	0.74	0.78	0.78	0.78

Table 3: *Classification performance.* Performance of Logistic Regression (LR) and RoBERTa (RB) on the two perspective-getting categories. P stands for precision, R for recall and \sum for macro-average. The results for RoBERTa are means from three runs with different seeds. Standard deviation is omitted as it stays consistently ≤ 0.03 . Best F1 performances are boldfaced.

dataset, AOP is present in 85% of the messages (n=1429), while AFC is only present in 25% of the messages (n=421). In total, we annotated 1,411 messages from the non-escalated and 263 from the escalated discussions.

5 Predicting Perspective-Getting

We now use our annotated dataset to train and test classification models to predict the two perspective-getting categories.

5.1 Classifiers

As a message can contain both perspective-getting categories, we train binary classifiers to predict the presence of each category separately. We experiment with a logistic regression model using a bag-of-words representation and a fine-tuned RoBERTa model. For the two perspective-getting categories, we divide the dataset into different training (64%), validation (16%) and test sets (20%). We do this to keep the same proportion of present and absent categories in the train, development and test sets.

We use logistic regression with L2 regularization as a baseline, implemented using scikit-learn (Pedregosa et al., 2011). Since Litvak et al. (2016) found that specific words can signal perspective-taking, we use the raw frequency of unigrams as features. We remove stop words using a list from the NLTK library (Bird and Loper, 2004). The punctuation signs {'?', '!', '.'} are treated as separate tokens and we replace all numbers with a <num> token. To penalize the model inversely proportional to the class frequency, we set the 'class weight' parameter to 'balanced'. We set the regularization parameter to 0.1, after experimenting with $\{10^{-3}, 10^{-2}, \dots, 10^2\}$ on the development set.

We fine-tune a RoBERTa model (Liu et al., 2019), with the default hyper-parameters: a learning rate of $5e^{-5}$, 500 warmup steps, and a dropout

	AOP		AFC
mention	0.41	?	1.22
word	0.37	wikipedia	-0.55
adding	-0.36	work	0.42
edit	-0.36	explain	0.38
know	0.35	feel	0.34

Table 4: *Logistic Regression Coefficients.* Features with the highest predictive weight derived from the Logistic Regression models for the Perspective-Getting Categories: Attending the Other’s Perspective (AOP) and Asking for Clarification (AFC).

probability of 0.1. We use the Adam optimizer, with $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1e^{-8}$ and a weight decay of 0.1. We use five training epochs based on results on the development set. We select the best performing model by evaluating on every 50 steps on the development set. We report the average results after fine-tuning on three random seeds.

5.2 Results

We evaluate the models based on precision, recall and F-1 scores. We report the macro average due to the uneven class distribution in our dataset, see Table 3. On average, RoBERTa outperforms the logistic regression model (see the F1 macro avg). RoBERTa performs worse for AOP (F1 of 0.74) than for AFC (F1 of 0.78). In contrast, logistic regression performs better for AOP (F1 of 0.83).

Feature and error analysis Table 4 shows highly-weighted features of our Logistic Regression model. Some of these features are specific to the Wikipedia-based training dataset, like *wikipedia* (AFC) and *edit* (AOP). The question mark, the most important feature for the AFC category, seems like a general sign of an attempt to clarify the other’s position. However, a question mark alone is not sufficient for predicting a genuine ask for clarification in many cases (18% of non-AFC utterances include a question mark, e.g., “Is this really difficult to understand for you??!” and 38% of AFC utterances do not include a question mark, e.g., Table 1).

For RoBERTa, we manually analyzed a sample of instances that the model predicted incorrectly (Table 5). In the first example, the statement “If there is no doubt it is a polish dialect” could be understood as the author expressing interest in the other’s perspective or a request to the previous author to explain their position. However, it could

Cat	Example	P	GT
AFC	<p>Message: There is no doubt it is a Polish dialect. One can doubt only by not knowing a history and listening to today's dialect. But go back to the beginning of the 20th century and read some texts written in the dialect from that period and you'll clearly see it is a Polish dialect with very little Czech influence. Today it sounds like a transitional dialect because of the presence (from 1920) of local people in the Czech state, which heavily linguistically influenced the dialect. But let's move some 500 metres to the Polish side of the border and you will hear more original version of the dialect (but more tainted by the correct Polish language). Btw. read pl:Dialekt śląski. -</p> <p>Reply: If there is no doubt it is a Polish dialect, I am sure you will find reliable sources to support the statement that it is a Polish dialect and not anything else. If you do not, the text should be altered. Before you support the statement by references or the article is altered, the NPOV tag should remain.</p>	0	1
AOP	<p>Message: NPOV</p> <p>Yes this is serious. The meanings of Abort Retry and Fail in this message were actually distinct. Abort would cause the program to return a TRUE value, fail would cause it to return a false value, and retry would simply retry the operation to see if something changed (like a disc being placed in a drive). Again, I'm totally serious about this concern.</p> <p>Reply: You're joking, right? If there's a factual inaccuracy in the article then state your reason here and make the change. I see no POV issues at all.</p>	1	0

Table 5: *Incorrect Model Predictions.* Examples of incorrectly classified messages on AFC and AOP categories, where the model predictions (P) differ from the annotation (GT) with high confidence.

also be read as an uptake of the previous position (i.e., AOP) without genuine interest in further explanation from the previous author. Similar to other areas in NLP, our task might be subject to plausible label variations (Pavlick and Kwiatkowski, 2019; Uma et al., 2021; Jiang and de Marneffe, 2022). However, overall, we reached reasonable agreement between annotators – note though that agreement was lower for AFC than for AOP.

For the second example, the model incorrectly predicted AOP. This prediction is understandable as it sounds like the responding author is paraphrasing the previous author's message by addressing the "factual inaccuracy" mentioned by them. However, when looking at the reply in context (i.e., in relation to the message it is answering to), it becomes clear that the responding author did not directly engage with anything specific the previous author said. When adding even more context, this becomes even clearer: The subsequent reply of the original author starts with "No, I'm sadly not joking. And no, it's not regarding factual inaccuracy [...]". Note that RoBERTa was only fine-tuned on the reply and did not see the original message. This might improve performance in future work.

6 How perspective-getting is associated with discussion course and outcomes

Perspective-getting may influence online discussions. First, conflict management and negotiation scholars found that exchanging information between negotiators positively influenced the negotiation process and outcomes (Trötschel et al.,

2011; Thompson and Hastie, 1990; Weingart et al., 1993; Adair and Brett, 2005). Second, political scientists demonstrated that exchanging narratives between in and out-group members in interpersonal conversation may reduce negative attitudes between the groups (Kalla and Broockman, 2020). Recently, Kalla and Broockman (2023) found that perspective-getting reduced negative stereotypes in inter-group interaction. In sum, there are indications that perspective-getting is an effective strategy to reduce conflict. We therefore now use our annotated dataset to study the influence of perspective-getting on online discussions' outcomes.

6.1 Hypotheses

Perspective-getting and escalations to a moderator In Wikipedia discussions, participants can request a moderator when they can not resolve a discussion by themselves (De Kock and Vlachos, 2021). Morrell (2010) suggest a relationship between interpersonal understanding and having constructive discussions without conflicts between citizens. Further, perspective-getting may reduce conflicts between opposing groups by decreasing thinking in stereotypes (Kalla and Broockman, 2023). Hence, our first hypothesis (H1) is that higher perspective-getting in a discussion are related to a lower rate of escalations to a moderator.

Initial perspective-getting and the course of the discussion Scholars have demonstrated how important the first message in an online conversation is for the course of the conversation (Zhang et al., 2018). Additionally, previous studies have found

that perspective-getting is related to increased positive intentions towards the other and their social group (Kalla and Broockman, 2023; Pagotto, 2010). As perspective-getting is considered as a more effective strategy to understand others (Damen et al., 2021; Eyal et al., 2018; Kalla and Broockman, 2023), we expect a similar influence on the course of the discussion: perspective-getting breeds perspective-getting. Therefore, our second hypothesis (H2) is that subsequent levels of perspective-getting in a discussion are positively related to higher perspective-getting in the first reply. Consequently, the null hypothesis states that the aforementioned relationship is insignificant.

6.2 Control Variables

The course and outcomes of discussions on Wikipedia do not only depend on perspective-getting behavior. There are several factors that need to be controlled for when testing hypotheses. First, discussions may differ in deliberation: how much effort discussants put in providing support for the arguments. Deliberation among discussants is known to decrease the likelihood of escalation. Second, discussions vary in the tone of how discussants communicate with each other. A disrespectful tone, e.g. with insults, is a typical cue for content moderation (Stockinger et al., 2023). Third, discussions may vary in how controversial the topic is. Controversial discussions on Wikipedia, e.g. about religion, are more likely to escalate and require moderation (Yasseri et al., 2014).

Providing Support of own Arguments (PSA)

We control for whether the discussants provide evidence for their own arguments while replying to the original message, which is a sign that they take the effort to justify their claims in a discussion. It is a key measure of deliberation in online discussions (Lundgaard and Etter, 2022), which may decrease inter-group conflict in communication. We annotated PSA along the same process as the perspective-getting categories. See an example in Table 6. Cohen's kappa was 0.86. In the annotated data, 38% (637) of the messages show PSA compared to 62% (1037) of the messages where PSA was absent.

Respectful Tone of Voice (RTV) We control for the extent to which discussants exercised a neutral to respectful tone of voice towards the other. Tone of voice has been suggested to be an important cue to recognize genuine interpersonal understanding,

as it signals a discussant's intention to engage with other person's perspective (Nilsen and Bacso, 2017; Keysar, 1994). See an example in Table 6. The Cohen's kappa was 0.75. In the annotated data, 79% (1325) of the messages show RTV.

Topic controversiality Studies have shown that how controversial a discussion topic is can influence discussion course and outcomes, such as escalation (Rad and Barbosa, 2012; Alashri et al., 2015). We therefore include how controversial a topic is as a control variable to our analyses. Previous research made a ranking of the most controversial topic categories on Wikipedia (Yasseri et al., 2014). This research identified and ranked controversial topic categories on Wikipedia: politics, countries, religion, history and social issues. We used this ranking to develop an ordinal scale to measure *topic controversiality*. This scale ranged from 5 for a discussion thread about the most controversial category (politics) to 0 for discussion threads about a non-controversial topic category (e.g., music). We developed a codebook for this scale to label the discussion threads in our dataset with the topic categories. For example, the Wikipedia discussion thread "List of sovereign states"⁴ was labeled with the countries category, and received a value of 4 as countries is the second most controversial topic category in the ranking (Yasseri et al., 2014). Two annotators used the codebook to label the topics of all discussion threads into topic categories (N=336). The intraclass correlation of a doubly coded random sample (N=69) was very good (0.77), implying a high level of agreement between annotators.

6.3 Results

Perspective-getting and escalations to a moderator (H1) We first conducted the Levene's test (Schultz, 1985) to test the t-test's equality of variance assumption for the perspective-getting of messages in escalated and non-escalated conversations. The test indicated equal variance ($F = 0.5$, $p = 0.82$). We then conducted a one-way ANCOVA to test our first hypothesis. ANCOVA (analysis of covariance) is a statistical method that adjusts for control variables to isolate the influence of categorical predictors (escalation) on an outcome variable (perspective-getting). The null hypothesis stated that the perspective-getting variables, AOP and

⁴https://en.wikipedia.org/wiki/List_of_sovereign_states

Category	Example
Respectful Tone of Voice (RTV)	✓ Message: The images are of poor quality and editorially meaningless. They will continue to be removed until better images or explanations can be found of their editorial utility. Reply: They are the best images out there. Please do not remove the pictures. They are all relevant to the article.
	✗ Message: I think you're lost. Democrat Underground is over here. Reply: [USERNAME], It's disgusting to know that there are people with a bellow 20 IQ like you on the internet. I really hope you're not a U.S. citizen. What a stain on this country you are!
Providing Support of own Arguments (PSA)	✓ Message: One of the problems with your proposal is that it never describes how popular Interstate Football was in Victoria. I am willing to change the line to (But Interstate Football mostly though out history was very popular in Victoria), to describe again that not always, but mostly though out history it was very popular in Victoria. But I won't support not describing what is a fact, and supported in many places that Interstate Football was mostly though out history very popular in Victoria. Reply: Read this article from 1927 and this article from 1944 .
	✗ Message: And how is that relevant to this article or this discussion? This is not a forum. Cheers, Reply: People die from censorship in China. Cheers as you put it.

Table 6: *Control Variables.* We display two example messages manifesting (✓) or not manifesting (✗) the considered control variable.

AFC, have higher scores for escalated messages than for non-escalated messages or there was no difference. Two ANCOVA tests were conducted to investigate the effect of AOP and AFC separately.

For AOP, the ANCOVA results did not reject the null hypothesis. While controlling for topic controversiality, PSA and RTV scores, there was no significant difference between non-escalated ($M = 0.86$, $SD = 0.35$, 95% CI [0.84, 0.87]) and escalated discussions ($M = 0.84$, $SD = 0.36$, 95% CI [0.8, 0.89]): $F(1,1668) = 0.12$, $p = 0.73$, $\eta_p^2 < 0.01$, $N = 1,674$. For AFC, the ANCOVA results did not reject the null hypothesis. There was also no significant difference between non-escalated ($M = 0.26$, $SD = 0.44$, 95% CI [0.24, 0.28]) and escalated discussions ($M = 0.21$, $SD = 0.41$, 95% CI [0.16, 0.26]): $F(1,1668) = 0.64$, $p = 0.42$, $\eta_p^2 < 0.01$, $N = 1,674$.

Our hypothesis is thus not supported: perspective-getting is not higher in non-escalated conversations than in escalated ones. This contrast with previous research on perspective-getting's positive influence on negative inter-group attitudes (Kalla and Broockman, 2023). First, perspective-getting may play less of a role because the goal on Wikipedia is to reach agreement (Damen et al., 2021; De Kock and Vlachos, 2021). Second, De Kock and Vlachos (2021) found that the best predictors to avoid escalation were politeness and collaboration. We controlled for RTV, a related concept to politeness, possibly leading to reduced effects. Last, we did not consider previous discussions between editors that may increase the need for third-party escalation.

Initial perspective-getting and the course of the discussion (H2)

For this hypothesis, only discussions where the first reply message disagreed with the opening post were included, resulting in 262 out of 399 annotated discussions. A multivariate regression analysis was conducted to test the hypothesis. The effect of AOP and AFC were tested separately, while controlling for topic controversiality. The regression analysis showed that the relationship between the AOP score of the first reply and the whole discussion is significant ($b = 0.9$, $t(259) = 20.48$, $p = 0.04$, 95% CI [0.06, 0.17]). Topic controversiality had no significant relationship with the AOP score ($b = -0.08$, $t(259) = -1.13$, $p = 0.26$, 95% CI [-0.24, 0.07]). The effect of AFC was not significant: $b = 0.03$, $t(259) = 1.69$, $p = 0.47$, 95% CI [-0.06, 0.12]). H2 is thus partially supported: the level of AOP in the discussion are positively associated with AOP in the first reply. This association was not found for AFC.

Since perspective-getting could have a diminishing influence over time, we tested the curvilinear association of the summated perspective-getting scores (sum of AOP and AFC) during the course of discussions. A hierarchical multiple regression analysis was conducted with the mean perspective-getting score of a discussion as a dependent variable, and perspective-getting scores of the first reply and topic controversiality as control variables. The hierarchical multiple regression revealed that the perspective-getting score of the first reply contributed significantly to the regression model, $F(1,260) = 9.99$, $p = 0.002$) and accounted

for 3.7% of the variation of the mean perspective-getting score of a discussion. Adding the quadratic polynomial to the model explained additional 0.4% of the variation, however, this change was not significant: $F(2, 259) = 5.61, p = 0.04$. Introducing the cubic polynomial explained 1% of additional variance, $F(3, 258) = 3.74, p = 0.01$, but this change also was not significant. Hence, we found evidence that perspective-getting in a first reply has a diminishing influence on perspective-getting in the discussion over time.

Although we measured perspective-getting instead of perspective-taking, these findings on H2 are partially in line with previous studies on the positive influence of empathetic reactions on harmonious communication (Pagotto, 2010) and the inhibiting influence on interpersonal aggression in discussions (Richardson et al., 1994).

7 Limitations

First, annotation was sometimes challenging due to characteristics of our data. Some replies were intended for a different user, such as the opening poster of the discussion, rather than the user who was directly replied to, which could affect the annotation of AOP. Furthermore, users sometimes added links to the text with the ‘insert link’ function. These links, however, are not visible in the dataset, which could affect the annotation of the PSA control variable, as links could lead to sources that do not support one’s position.

Second, perspective-getting can be operationalized in a more fine-grained way than our binary categories. For example, just acknowledging a perspective (e.g., “I understand that your view is X. But here is what I think ...”) could be seen as less AOP than actually engaging with it (e.g., “It seems the point you are making is X. I understand how this might come from the belief that ...”).

Third, the number of annotated messages is relatively low. Due to the complexity of perspective-getting, we opted for careful annotation by expert annotators instead of item quantity. Due to the small number of escalated discussions in the original dataset, the number of escalated discussions is especially small (c.f. Table 2), possibly influencing the generalizability of the testing of hypothesis 1.

Fourth, the generalizability of our study is limited, as perspective-getting might manifest differently in other domains. For example, Wikipedia editors have to come to an agreement, while so-

cial media users discussing presidential candidates do not. Further, most replies in our dataset will in some form disagree with the previous message, including those with high perspective-getting. Still, our codebook was developed based on theoretical work and should be transferable to other settings, possibly with some minor changes.

Fifth, our data sample is in English and is likely skewed to specific demographics. Our classifiers are therefore likely to be biased towards these same demographics in performance. Furthermore, our classifiers could latch on to spurious features. We do not recommend using the classifier “out of the box” on datasets that are not comparable.

8 Conclusion

In this study, we measured perspective-getting in Wikipedia discussions. Based on theory, we developed a codebook to annotate messages in Wikipedia discussions on two perspective-getting categories, and then fine-tuned a RoBERTa model to predict the categories. In our dataset, perspective-getting was not higher in non-escalated discussions. However, discussions starting with a post attending the other’s perspective are more often followed by responses that also attend to other’s perspective.

With this work, we hope to contribute to studies of how and when strategies for interpersonal understanding can reduce polarization in online discussions (Kwon and Cho, 2017; Shmargad et al., 2022), as perspective-getting allows people to understand, consider and adopt the opposite position in discussions (Eyal et al., 2018; Kalla and Broockman, 2023; Damen et al., 2021). Future research could build on our work to study the relation between perspective-getting and online polarization, while considering the contextual and causal limitations of our study.

Acknowledgements

We thank the anonymous reviewers. This research was supported by the “Digital Society - The Informed Citizen” research programme, which is (partly) financed by the Dutch Research Council (NWO), project 410.19.007.

References

Wendi L Adair and Jeanne M Brett. 2005. The negotiation dance: Time, culture, and behavioral sequences in negotiation. *Organization Science*, 16(1):33–51.

- Saud Alashri, Sultan Alzahrani, Lenka Bustikova, David Siroky, and Hasan Davulcu. 2015. What animates political debates? Analyzing ideological perspectives in online debates between opposing parties. In *Proceedings of the ASE/IEEE International Conference on Social Computing (SocialCom-15)*.
- Ofer Arazy, Lisa Yeo, and Oded Nov. 2013. Stay on the Wikipedia task: When task-related disagreements slip into personal and procedural conflicts. *Journal of the American Society for Information Science and Technology*, 64(8):1634–1648.
- Carolyn M Axtell, Sharon K Parker, David Holman, and Peter Totterdell. 2007. Enhancing customer service: Perspective taking in a call centre. *European Journal of Work and Organizational Psychology*, 16(2):141–168.
- Valentin Barriere, Shabnam Tafreshi, João Sedoc, and Sawsan Alqahtani. 2022. WASSA 2022 shared task: Predicting empathy, emotion and personality in reaction to news stories. In *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, pages 214–227, Dublin, Ireland. Association for Computational Linguistics.
- Steven Bird and Edward Loper. 2004. NLTK: The natural language toolkit. In *Proceedings of the ACL Interactive Poster and Demonstration Sessions*, pages 214–217, Barcelona, Spain. Association for Computational Linguistics.
- Laura W Black, Howard T Welser, Dan Cosley, and Jocelyn M DeGroot. 2011. Self-governance through group discussion in Wikipedia: Measuring deliberation in online groups. *Small Group Research*, 42(5):595–634.
- Emile G Bruneau and Rebecca Saxe. 2012. The power of being heard: The benefits of ‘perspective-giving’ in the context of intergroup conflict. *Journal of experimental social psychology*, 48(4):855–866.
- Sven Buechel, Anneke Buffone, Barry Slaff, Lyle Ungar, and João Sedoc. 2018. Modeling empathy and distress in reaction to news stories. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4758–4765, Brussels, Belgium. Association for Computational Linguistics.
- Herbert H Clark. 1996. *Using language*. Cambridge university press.
- Herbert H Clark and Susan E Brennan. 1991. Grounding in communication.
- Debby Damen, Monique MH Pollmann, and Teri-Louise Grassow. 2021. The benefits and obstacles to perspective getting. *Frontiers in Communication*, 6:611187.
- Debby Damen, Marije van Amelsvoort, Per van der Wijst, and Emiel Krahmer. 2019. Changing views: the effect of explicit perception-focus instructions on perspective-taking. *Journal of Cognitive Psychology*, 31(3):353–369.
- Debby Damen, Per van der Wijst, Marije van Amelsvoort, and Emiel Krahmer. 2020. Can the curse of knowing be lifted? the influence of explicit perspective-focus instructions on readers’ perspective-taking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(8):1407.
- Christine De Kock and Andreas Vlachos. 2021. I beg to differ: A study of constructive disagreement in online conversations. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2017–2027, Online. Association for Computational Linguistics.
- Christine De Kock and Andreas Vlachos. 2022. How to disagree well: Investigating the dispute tactics used on Wikipedia. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 3824–3837, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- Bill Dolan and Chris Brockett. 2005. Automatically constructing a corpus of sentential paraphrases. In *Third International Workshop on Paraphrasing (IWP2005)*.
- Wafaa S. El-Kassas, Cherif R. Salama, Ahmed A. Rafea, and Hoda K. Mohamed. 2021. Automatic text summarization: A comprehensive survey. *Expert Systems with Applications*, 165:113679.
- Nicholas Epley, Boaz Keysar, Leaf Van Boven, and Thomas Gilovich. 2004a. Perspective taking as egocentric anchoring and adjustment. *Journal of personality and social psychology*, 87(3):327.
- Nicholas Epley, Carey K Morewedge, and Boaz Keysar. 2004b. Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of experimental social psychology*, 40(6):760–768.
- Tal Eyal, Mary Steffel, and Nicholas Epley. 2018. Perspective mistaking: Accurately understanding the mind of another requires getting perspective, not taking perspective. *Journal of personality and social psychology*, 114(4):547.
- Adam D. Galinsky, William W. Maddux, Debra Gilin, and Judith B. White. 2008. Why it pays to get inside the head of your opponent: The differential effects of perspective taking and empathy in negotiations. *Psychological Science*, 19(4):378–384. Publisher: SAGE Publications Inc.
- James Gibson, Nikolaos Malandrakis, Francisco Romero, David C. Atkins, and Shrikanth S.

- Narayanan. 2015. Predicting therapist empathy in motivational interviews using language features inspired by psycholinguistic norms. In *Proc. Inter-speech 2015*, pages 1947–1951.
- Yun He, Zhuoer Wang, Yin Zhang, Ruihong Huang, and James Caverlee. 2020. PARADE: A new dataset for paraphrase identification requiring computer science domain knowledge. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7572–7582, Online. Association for Computational Linguistics.
- Claire Hughes and Sue Leekam. 2004. What are the links between theory of mind and social relations? review, reflections and new directions for studies of typical and atypical development. *Social development*, 13(4):590–619.
- Nan-Jiang Jiang and Marie-Catherine de Marneffe. 2022. Investigating reasons for disagreement in natural language inference. *Transactions of the Association for Computational Linguistics*, 10:1357–1374.
- Joshua L Kalla and David E Broockman. 2020. Reducing exclusionary attitudes through interpersonal conversation: Evidence from three field experiments. *American Political Science Review*, 114(2):410–425.
- Joshua L Kalla and David E Broockman. 2023. Which narrative strategies durably reduce prejudice? Evidence from field and survey experiments supporting the efficacy of perspective-getting. *American Journal of Political Science*, 67(1):185–204.
- Aikaterini-Lida Kalouli, Rebecca Kehlbeck, Rita Sevastjanova, Oliver Deussen, Daniel Keim, and Miriam Butt. 2021. Is that really a question? Going beyond factoid questions in NLP. In *Proceedings of the 14th International Conference on Computational Semantics (IWCS)*, pages 132–143, Groningen, The Netherlands (online). Association for Computational Linguistics.
- Jenna Kanerva, Filip Ginter, Li-Hsin Chang, Iiro Rastas, Valtteri Skantsi, Jemina Kilpeläinen, Hanna-Mari Kupari, Aurora Piirto, Jenna Saarni, Maija Sevón, et al. 2023. Towards diverse and contextually anchored paraphrase modeling: A dataset and baselines for Finnish. *Natural Language Engineering*, pages 1–35.
- B. Keysar. 1994. The illusory transparency of intention: Linguistic perspective taking in text. *Cognitive Psychology*, 26(2):165–208.
- Boaz Keysar, Shuhong Lin, and Dale J Barr. 2003. Limits on theory of mind use in adults. *Cognition*, 89(1):25–41.
- Hamed Khanpour, Cornelia Caragea, and Prakhar Biyani. 2017. Identifying empathetic messages in online health communities. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 246–251, Taipei, Taiwan. Asian Federation of Natural Language Processing.
- Robert M Krauss and Susan R Fussell. 1991. Perspective-taking in communication: Representations of others’ knowledge in reference. *Social cognition*, 9(1):2–24.
- K Hazel Kwon and Daegon Cho. 2017. Swearing effects on citizen-to-citizen commenting online: A large-scale exploration of political versus nonpolitical online news sites. *Social Science Computer Review*, 35(1):84–102.
- Allison Lahnala, Charles Welch, David Jurgens, and Lucie Flek. 2022. A critical reflection and forward perspective on empathy and natural language processing. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 2139–2158, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- J. Richard Landis and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.
- Marina Litvak, Jahna Otterbacher, Chee Siang Ang, and David Atkins. 2016. Social and linguistic behavior and its correlation to trait empathy. In *Proceedings of the Workshop on Computational Modeling of People’s Opinions, Personality, and Emotions in Social Media (PEOPLES)*, pages 128–137, Osaka, Japan. The COLING 2016 Organizing Committee.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv:1907.11692 [cs]*.
- Daniel Lundgaard and Michael Etter. 2022. Everyday talk on Twitter: Informal deliberation about (ir-) responsible business conduct in social media arenas. *Business & Society*, 62(6):1201–1247.
- Michael E Morrell. 2010. *Empathy and democracy: Feeling, thinking, and deliberation*. Penn State Press.
- Elizabeth S. Nilsen and Sarah A. Bacso. 2017. Cognitive and behavioural predictors of adolescents’ communicative perspective-taking and social relationships. *Journal of Adolescence*, 56:52–63.
- Lisa Pagotto. 2010. *The interplay of empathy, oneness and perceived similarity in mediating the effects of perspective taking on prosocial responses*. Ph.D. thesis, Università degli Studi di Padova.
- Silvia Pareti, Tim O’Keefe, Ioannis Konstas, James R. Curran, and Irena Koprinska. 2013. Automatically detecting and attributing indirect quotations. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 989–999, Seattle, Washington, USA. Association for Computational Linguistics.

- Ellie Pavlick and Tom Kwiatkowski. 2019. Inherent Disagreements in Human Textual Inferences. *Transactions of the Association for Computational Linguistics*, 7:677–694.
- 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.
- Hoda Sepehri Rad and Denilson Barbosa. 2012. Identifying controversial articles in wikipedia: A comparative study. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*, WikiSym '12, New York, NY, USA. Association for Computing Machinery.
- Sudha Rao and Hal Daumé III. 2018. Learning to ask good questions: Ranking clarification questions using neural expected value of perfect information. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2737–2746, Melbourne, Australia. Association for Computational Linguistics.
- Sudha Rao and Hal Daumé III. 2019. Answer-based adversarial training for generating clarification questions. 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 143–155, Minneapolis, Minnesota. Association for Computational Linguistics.
- Anu Realo, Jüri Allik, Aire Nolkav, Raivo Valk, Tuuli Ruus, Monika Schmidt, and Tiina Eilola. 2003. Mind-reading ability: Beliefs and performance. *Journal of Research in Personality*, 37(5):420–445.
- Deborah R. Richardson, Georgina S. Hammock, Stephen M. Smith, Wendi Gardner, and Manuel Signo. 1994. Empathy as a cognitive inhibitor of interpersonal aggression. *Aggressive Behavior*, 20(4):275–289.
- Carl Rogers and Richard Farson. 1957. *Active listening*. Mockingbird Press LLC.
- Tulika Saha, Aditya Prakash Patra, Sriparna Saha, and Pushpak Bhattacharyya. 2020. A transformer based approach for identification of tweet acts. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.
- Tulika Saha, Apoorva Upadhyaya, Sriparna Saha, and Pushpak Bhattacharyya. 2021. Towards sentiment and emotion aided multi-modal speech act classification in Twitter. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5727–5737, Online. Association for Computational Linguistics.
- Brian B Schultz. 1985. Levene’s test for relative variation. *Systematic Zoology*, 34(4):449–456.
- Ashish Sharma, Adam Miner, David Atkins, and Tim Althoff. 2020. A computational approach to understanding empathy expressed in text-based mental health support. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5263–5276, Online. Association for Computational Linguistics.
- Yotam Shmargad, Kevin Coe, Kate Kenski, and Stephen A Rains. 2022. Social norms and the dynamics of online incivility. *Social Science Computer Review*, 40(3):717–735.
- Andrea Stockinger, Svenja Schäfer, and Sophie Lecheler. 2023. Navigating the gray areas of content moderation: Professional moderators’ perspectives on uncivil user comments and the role of (ai-based) technological tools. *new media & society*, page 14614448231190901.
- Leigh Thompson and Reid Hastie. 1990. Social perception in negotiation. *Organizational behavior and human decision processes*, 47(1):98–123.
- Roman Trötschel, Joachim Hüffmeier, David D Loschelder, Katja Schwartz, and Peter M Gollwitzer. 2011. Perspective taking as a means to overcome motivational barriers in negotiations: When putting oneself into the opponent’s shoes helps to walk toward agreements. *Journal of personality and social psychology*, 101(4):771.
- Don Tugener, Margot Mieskes, Jan Deriu, and Mark Cieliebak. 2021. Are we summarizing the right way? A survey of dialogue summarization data sets. In *Proceedings of the Third Workshop on New Frontiers in Summarization*, pages 107–118, Online and in Dominican Republic. Association for Computational Linguistics.
- Alexandra N Uma, Tommaso Fornaciari, Dirk Hovy, Silviu Paun, Barbara Plank, and Massimo Poesio. 2021. Learning from disagreement: A survey. *Journal of Artificial Intelligence Research*, 72:1385–1470.
- Kai Wang and Tat-Seng Chua. 2010. Exploiting salient patterns for question detection and question retrieval in community-based question answering. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 1155–1163, Beijing, China.
- Laurie R Weingart, Rebecca J Bennett, and Jeanne M Brett. 1993. The impact of consideration of issues and motivational orientation on group negotiation process and outcome. *Journal of Applied Psychology*, 78(3):504.
- Bo Xiao, Dogan Can, Panayiotis G. Georgiou, David Atkins, and Shrikanth S. Narayanan. 2012. Analyzing the language of therapist empathy in motivational interview based psychotherapy. In *Proceedings of The 2012 Asia Pacific Signal and Information*

Processing Association Annual Summit and Conference, volume 2012, pages 1–4.

Jingjing Xu, Yuechen Wang, Duyu Tang, Nan Duan, Pengcheng Yang, Qi Zeng, Ming Zhou, and Xu Sun. 2019. Asking clarification questions in knowledge-based question answering. 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 1618–1629, Hong Kong, China. Association for Computational Linguistics.

Taha Yasseri, Anselm Spierri, Mark Graham, and János Kertész. 2014. *The most controversial topics in Wikipedia: A multilingual and geographical analysis*, volume 25. Scarecrow Press.

Hamed Zamani, Susan Dumais, Nick Craswell, Paul Bennett, and Gord Lueck. 2020. Generating clarifying questions for information retrieval. In *Proceedings of The Web Conference 2020, WWW '20*, page 418–428, New York, NY, USA. Association for Computing Machinery.

Justine Zhang, Jonathan Chang, Cristian Danescu-Niculescu-Mizil, Lucas Dixon, Yiqing Hua, Dario Taraborelli, and Nithum Thain. 2018. Conversations gone awry: Detecting early signs of conversational failure. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1350–1361, Melbourne, Australia. Association for Computational Linguistics.

Ke Zhou, Luca Maria Aiello, Sanja Scepanovic, Daniele Quercia, and Sara Konrath. 2021. The language of situational empathy. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW1).

A Codebook

HOW TO CODE 1 = the utterance fits with the perspective-getting category (present) 2 = the utterance does not fit with the perspective-getting category (absent). To calculate the perspective-getting score of an utterance: $\#(\text{present})/\#(\text{absent or present})$

DEFINITIONS: Original message: message, which is replied by the current utterance; Current message: the message for which perspective-getting is measured; Author: a user who writes the reply, for which a perspective-getting score is evaluated.

REQUIREMENTS for an utterance pair labeling: 1. The interaction has to be between two separate users. If an interaction is a self-reply, its perspective-getting cannot be assessed and thus the utterance has to be skipped. 2. The interaction must contain a disagreement. If disputers fully agree with each other, the utterance pair should be skipped.

Category & Definition	Present		Absent	
	Description	Example	Description	Example
1. Attending the other's Perspective (AOP). Consideration of the position of an opponent is a crucial component of perspective-taking. For this criterion, it is necessary to ensure that the opponent's position or arguments are attended. An opposite behaviour would be to build own arguments without paying attention to the argumentation of an opponent. The adaptation of the position is not vital for this criterion.	Quoting or rephrasing the original message; Mentioning words/phrases/arguments in the original messages, while trying to argue with them; The answer involves constructive criticism or agreement with the ideas from the original message; Shifting own perspective under the influence of argumentation in the original message; Using phrases such as "I agree with you on...", "You are right about...", etc.	B: ...Technically spoken, Evropariver did right. He cited a source correctly and gave the references. Thus, in my opinion he did nothing wrong there. On the other hand: If someone claims things such as "Western biology begins in Egypt", my alert bells start to ring very loudly. Egypt is not "western"... Second, the author of the book Evropariver used claims that "Five thousand years ago Egyptian priests were already starting to gather a tremendous amount of medical data"...	The arguments or ideas from the original message are ignored. The current message only provides own argumentation. Switching from on-topic to off-topic (e.g., from the page discussion to interpersonal relations)	A: (Crimea is a part of Ukraine) De jure according to Ukrainian law only. No other country's law considers Crimea as Ukrainian land. B: Oh look, another editor with a tiny edit count who is totally not trying to minimize international recognition of Ukraine's territory. (facepalm).
2. Asking for Clarification (AFC). The author is interested in understanding of the alternative position and tries to obtain additional information about the alternative perspective.	Asking clarifying questions; Using phrases such as "If I understood you correctly...", "I think you meant..."	A: Actually, Mein Kampf contained quite a bit of criticism of Judaism: biased, self-serving and wrong-headed, of course, but criticism nonetheless. There's no particular reason that a person cannot be included in both categories which I think is the case here. B: Did he criticize Judaism as a religion or just the Jews as a group? Only in the first case, the category applies.	The position of the opponent is not clarified; Explanations of WikiDisputes rules would not be enough to satisfy this criteria as "present"	A: That's what I wrote according to the source, but it says that real story "Western biology begins in Egypt", please read more carefully when deleting. Also, why did you remove the invention of medicin attributed to Egypt and everything related to it, explain yourself. B: Did you read what I wrote above? Don't pretend you didn't see that. Your source contradicts itself. And why did you remove Lis Magner? Because you don't like it? That's not how we do things around here. And in case you hadn't noticed, this article is about biology, not medicine.
Control variable: Providing Support of own Arguments (PSA). Ensuring that the arguments will be understood by the opponent and acknowledging that the opponent might not know some information.	A person has put efforts to get across own ideas to persuade the opponent; Links to valuable sources are provided to get across own ideas more effectively; Explaining concepts or definitions that could be unknown or sound ambivalent for an opponent; Specific examples or facts instead of abstract ideas are provided; References to the WikiDisputes rules.	A: It's mentioned that psychology is a science in the first sentence, I don't understand the problem. B: I am not saying psychology is not a science (even though that's what I think), No I mean a section detected to the debate on whether psychology is a science or not. I can't access your link, but if you read Science it defines science as "a systematic enterprise that builds and organizes knowledge in the form of testable explanations and predictions about the universe". That describes psychology as well as any other science. Psychology uses the scientific method for much of its research, and has done so for many decades.	Arguments are provided in an abstract manner or not provided at all; People use concepts, phrases or abbreviations that were not mentioned in the original message, without explanations; No attempts to explain or support own position. Promises of evidences in subsequent messages would be considered as "absent" for this criterion Referencing to own experience and nothing else is considered as "absent"	A: can you please explain your concerns about this edit? I was undoing disruptive, autobiographical edits by a sockpuppet IP account. Some of the added content was unduly promotional or supported only by unreliable sources. B: You just deleted everything that someone did saying that it's all shit. I guess, YOU should explain your very aggressive behavior. Do you have any personal issues with the article or the person?
Control variable: Respectful Tone of Voice (RTV). The author uses neutral or respectful language. Exercising offensive or sarcastic language is usually associated with intolerance towards the alternative position.	The utterance has a neutral or respectful tone; No direct signs of hostility of disregard towards the alternative position are shown; Using such phrases as 'please', 'thank you', 'excuse me', etc. in a non-sarcastic manner.	B: I haven't read "Young Stalin" yet. I do not doubt your quote (and there was another source too,anyways) so this isn't any other than asking for a little help (Wikipedia aside), if you please just could tell me in which chapter (number or name) the English pages 310-311 are found, I would be very grateful.	Using hostile, offensive words Using sarcasm Making jokes about the opponent's perspective Reduction of the opponent's position to absurdity.	A: Except I compared the rankings with the whole of Europe, did you read what I said? And for some reason, you want to ignore Africa, Asia and Oceania And please assume good faith. B: ... You don't see anything written in front of you. North America and Latin America have nothing to do with Russia, absolutely nothing. What are you talking about? We'll compare a country to which continent it lies on, not to foreign continents, for example Africa. All this time, you did not contribute a single thing to the discussion, and just kept complaining. I'd suggest you to leave. Not a single thing sounds "sweet" in this sentence.

Table 7: **Annotation rules.** For each developed perspective-getting category, we display the description and example(s) message(s) when the considered perspective-getting category is absent or present. Note: the category names were changed after the annotation procedure for clarity. The initial categories' names were Considering Opponent Perspective (COP) instead of Attending the Other's Perspective (AOP), and Finding Common Ground instead of Asking for Clarification (AFC). As we focus on how interlocutors communicate to acquire more information about the other's perspective, we changed the initial AOP and AFC labels to better embed our research in perspective-getting instead of perspective-taking literature. The instructions of the codebook remained unchanged. Respectful Tone of Voice (RTV) and providing Support of own Arguments (PSA) remained the same. RTV and PSA were initially dimensions of perspective-taking in our study, but were excluded from the perspective-getting dimensions to increase constructive validity. Notations *A* and *B* in the *Example* columns represent original (those that were replied to) and current (the one we label) messages respectively.