

In search of civility online

What does civil discourse look like on the Internet and where can it be found?

We investigated these questions by examining 1) the Reddit r/ChangeMyView community and 2) instances where someone saw the need to intervene to turn an uncivil discussion into a civil one.

CMV user • 5y ago

Yes, or as someone else has said, **stop arguing** emotionally in this discussion. You can feel whatever, but this is a place for civil discourse (and I should know, since I have had posts removed because of this).



Team Leaders: Eetu Mäkelä,¹ Ümit Bedretdin¹
Participants: Ylva Biri,¹ Yichen Du,² Hibiki Ito,¹ Nan Jiang,¹ Kärt Kaal,¹ Charlotte Panušková,³ Marina Pavlova,⁴ Niko Petjakko,¹ Heljä Räisänen,¹ Natchanun Sanitdee,¹ Samuli Veikkolainen¹



Defining civility

Respect of different views and others - orientation toward understanding, not agreement.

Openness to discussion - seeking mutually acceptable outcomes.

Adherence to contexts - willingness to follow socially approved rules.

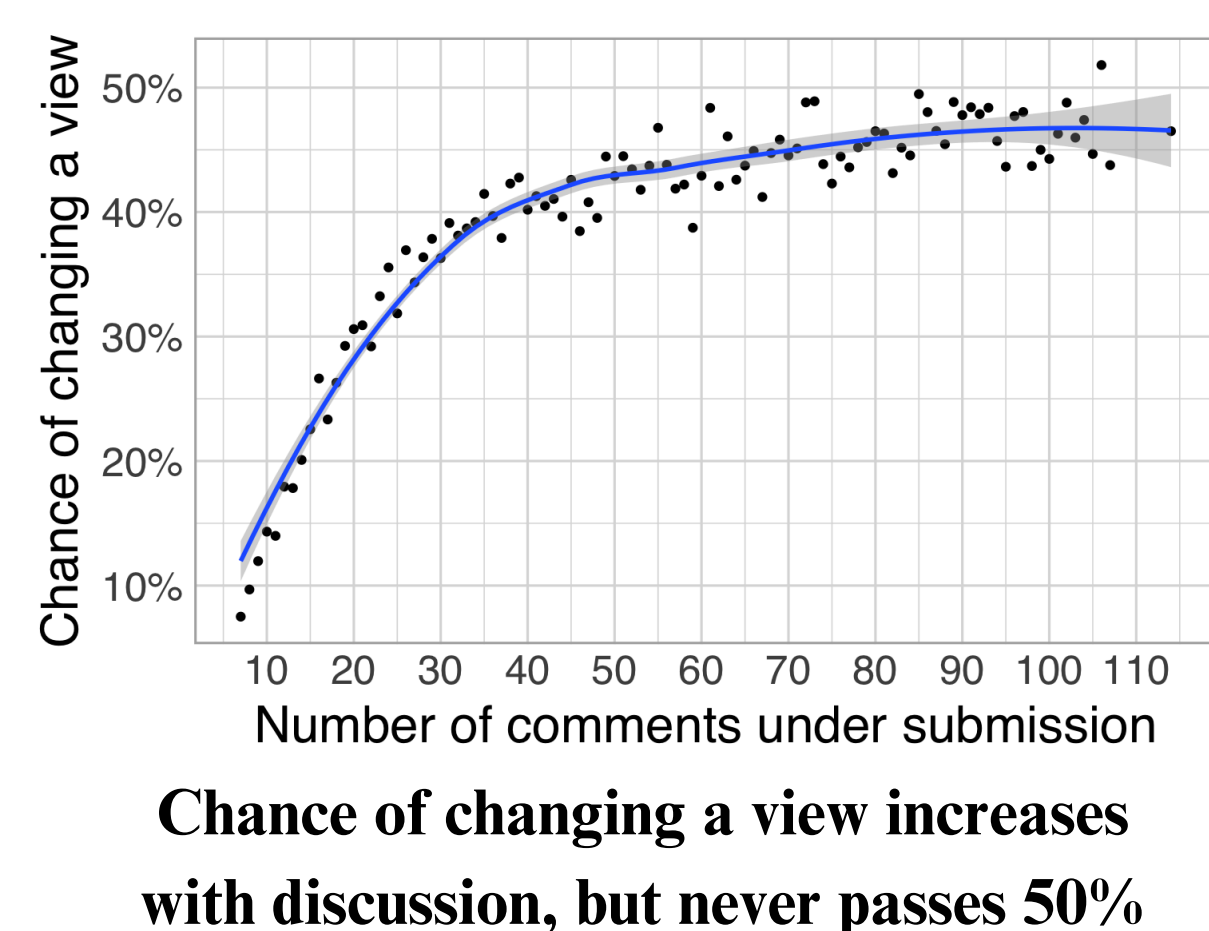
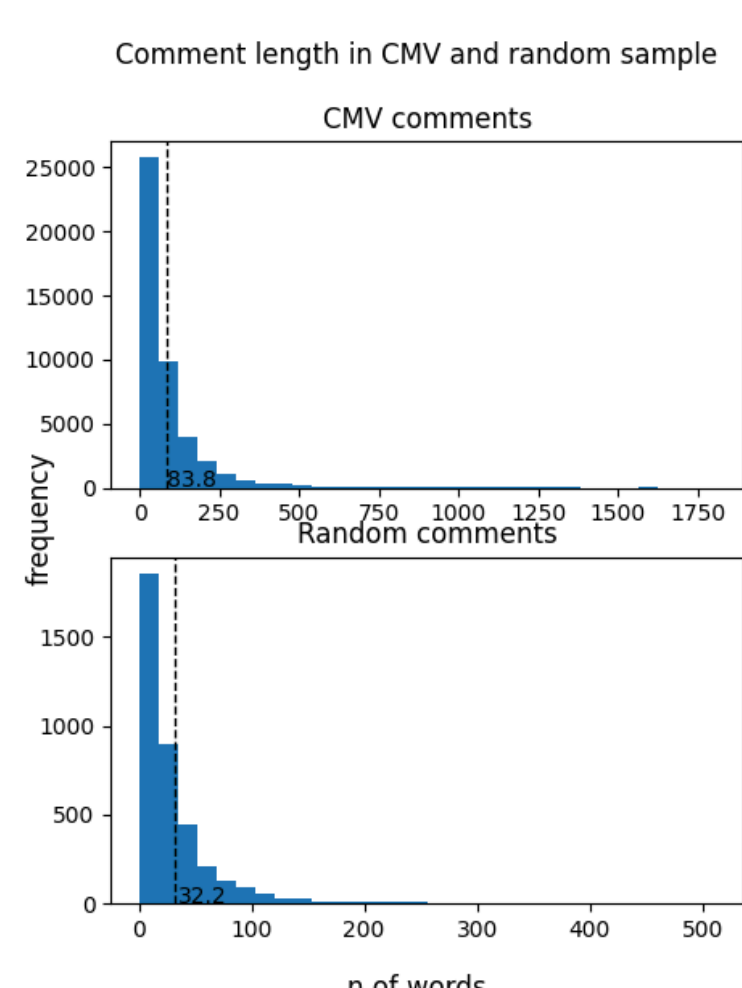
Staying on topic - civil discourse has “an eye toward...relevance”.

(Hauser, 1999; Kingwell, 1995; Massaro & Stryker, 2012; Papacharissi, 2004)

► The rules of the Reddit r/ChangeMyView community match these principles, making it a good basis for finding civil discourse online.

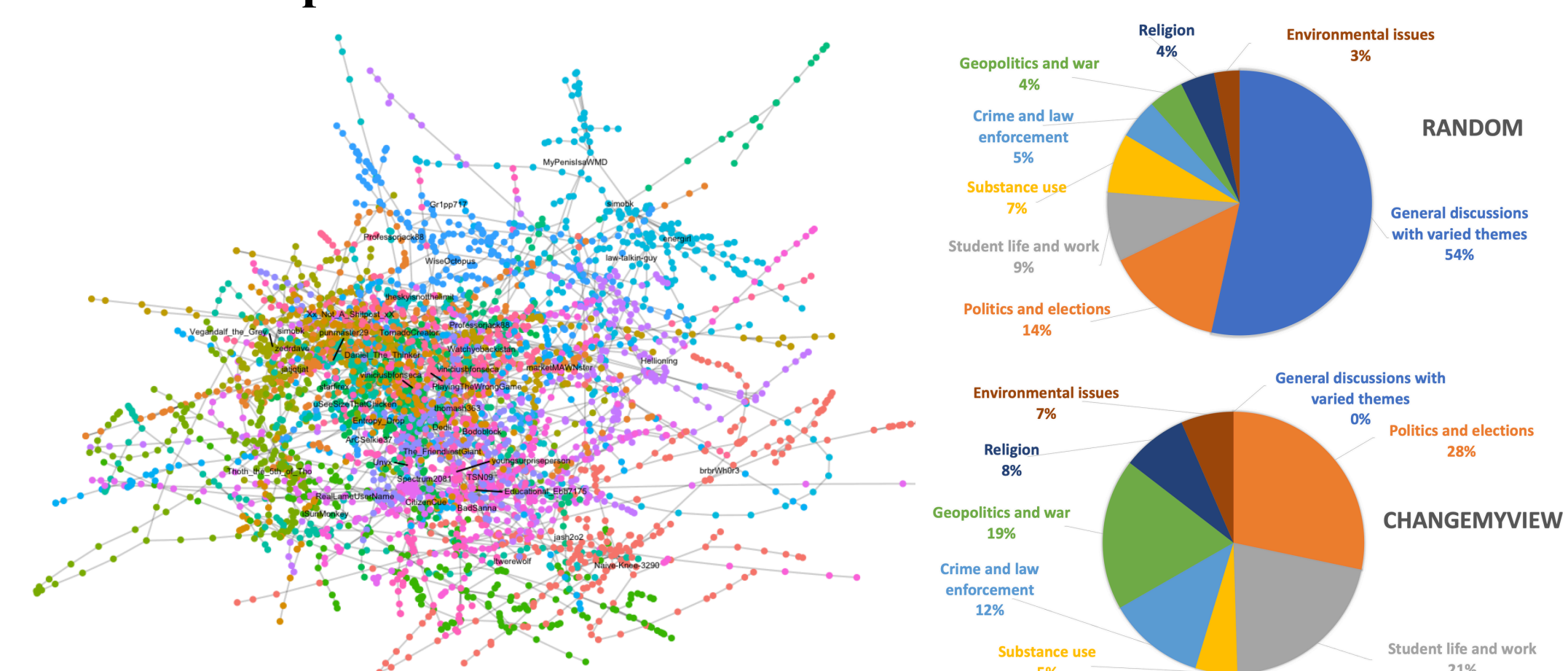
CMV as a community

- Centered around challenging each other's views.
- Rules enforce respectful and impactful discussion.
- Users give deltas (Δ) to comments that changed their mind.



Top 20 communities of comments and replies in CMV

Popular Topics

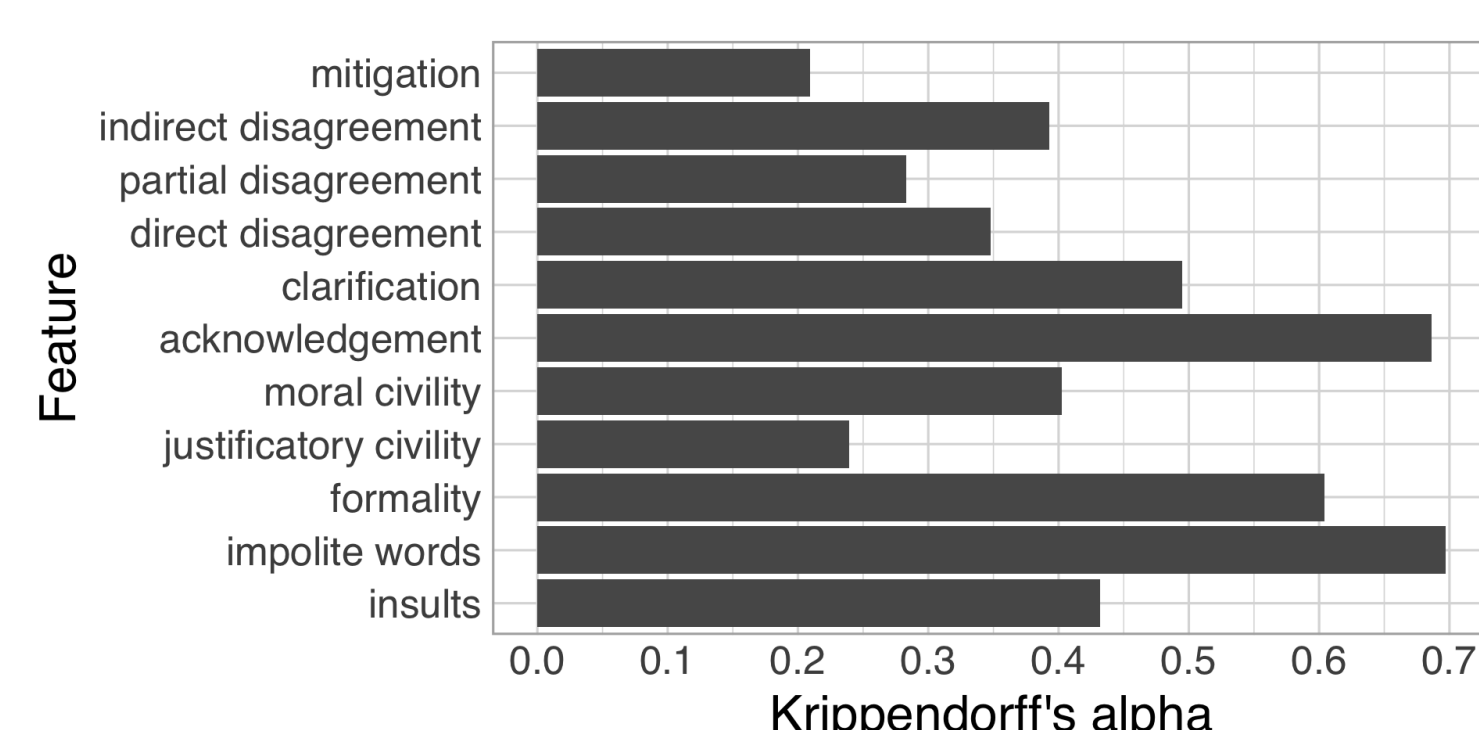


Close-reading civility on Reddit

Annotation Scheme

Initial Identification	Category	Sub-category	Sub-subcategory
Civility	Politeness	Mitigation (MD)	Correct me if I'm wrong...
		Indirect disagreement (ID)	What about ...?
		Partial disagreement (PD)	You're right but...
	Public-mindedness	Degree Formality (FM)	
		Acknowledgement (AK)	That's a great point.
		Clarification (CT)	Do you mean that ...?
Incivility	Impoliteness	Moral Civility (MC)	respecting fundamental rights
		Justificatory Civility (JC)	refer to law/objective rules, balance of values
Neutral	Direct (total) disagreement (DD)	Impolite Word (IM)	
		Insults (IS)	

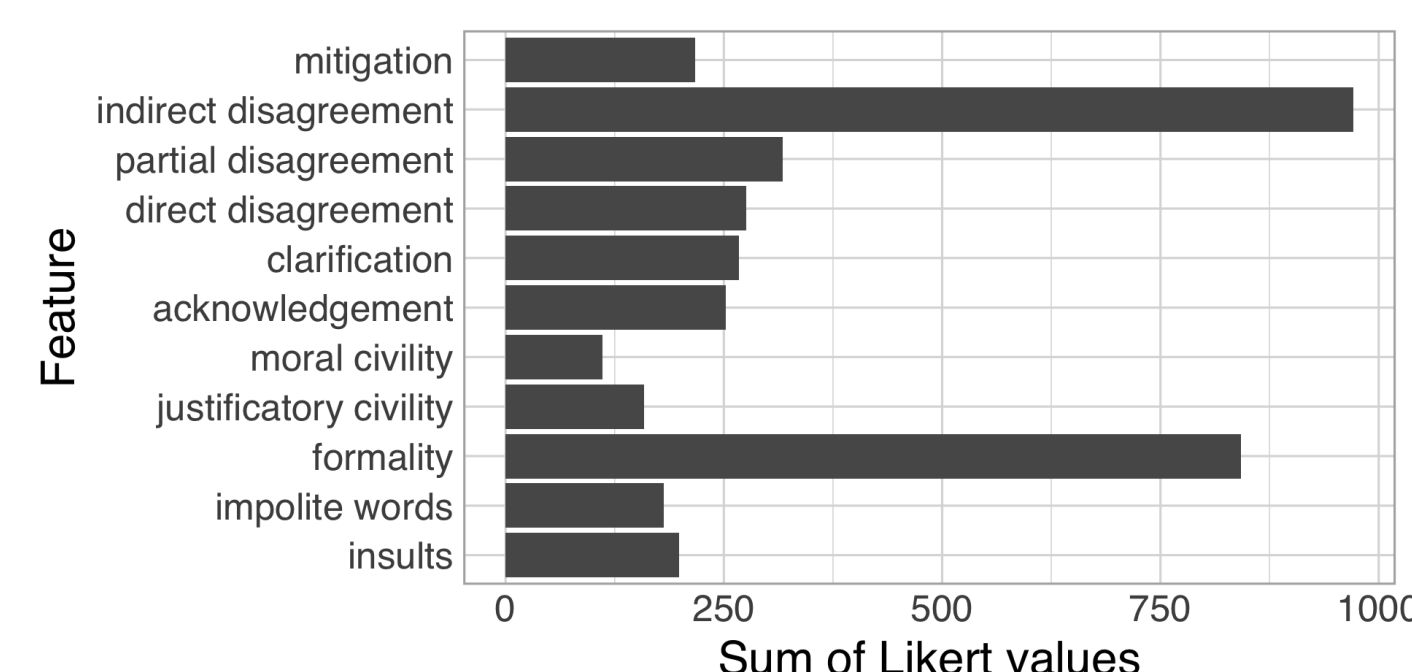
Each of the 11 features annotated on a 1-7 Likert scale.



Inter-annotator agreement by civility features

- Disagreement strategies are challenging to distinguish.
- Public-mindedness appears in longer comments, which are rare online. When detected, it is hard to rate how much is present.
- Agreement on insults suffers due to difficulties judging sarcasm.
- For downstream use, annotations collapsed into a binary civility indicator based on feature combinations (also alleviating uncertainty caused by annotation disagreements).

Theory meets data



Apart from indirect disagreement and formality, many of the civility features turned out to be rare in online environments.

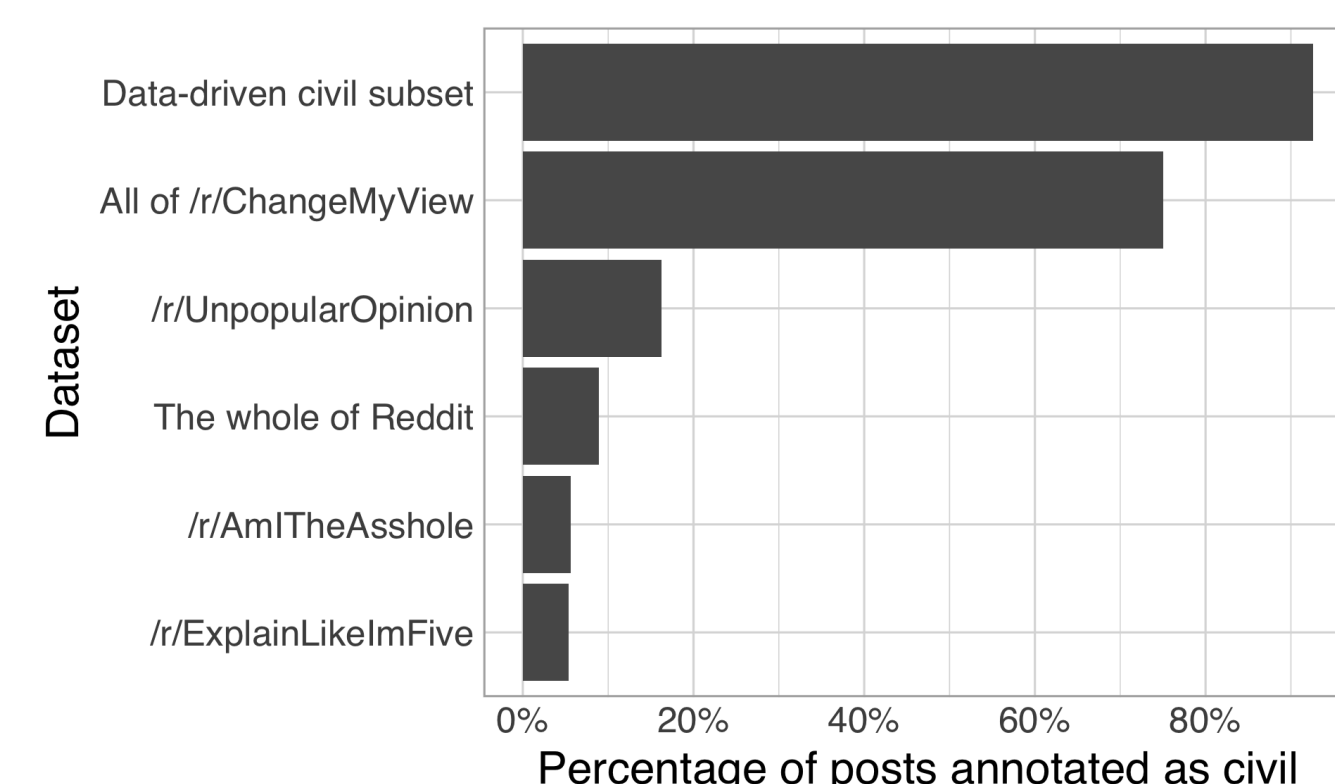
Extracting an exemplary civil discourse subset from CMV

Criteria:

- Comment needs to be in a delta-awarding thread (persuasive discussion likely to be civil).
- There needs to be at least one back-and-forth exchange before the delta is awarded (requirement of discussion).
- No parts of the discussion can have more downvotes than upvotes (filter out potential incivility).
- Comment needs to be less than 10 deep in a thread (through exploration, it was determined that long threads deteriorate).

► Even this quite strict filtering still leaves 164,592 comments as potential training data if the set turns out to be good!

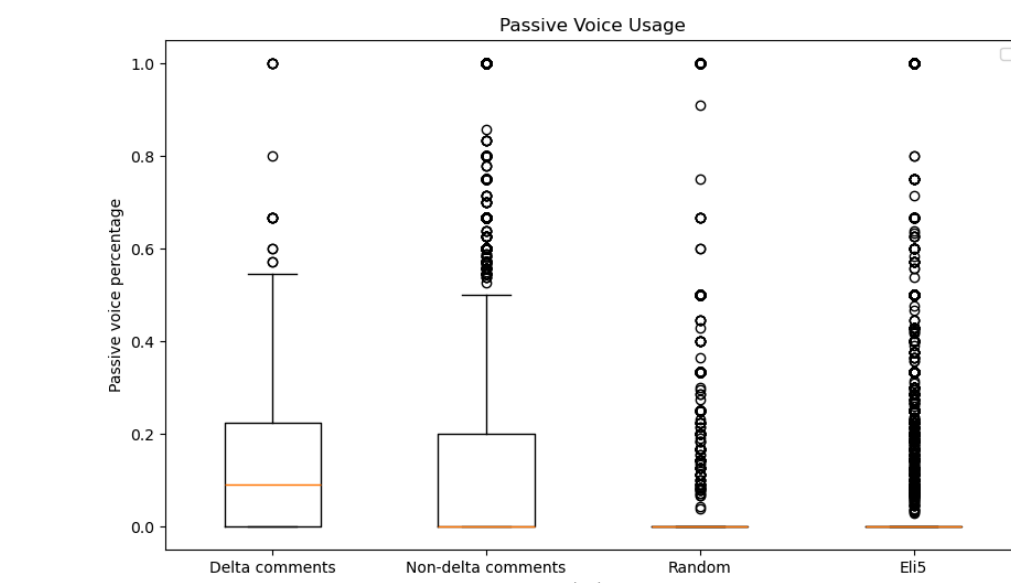
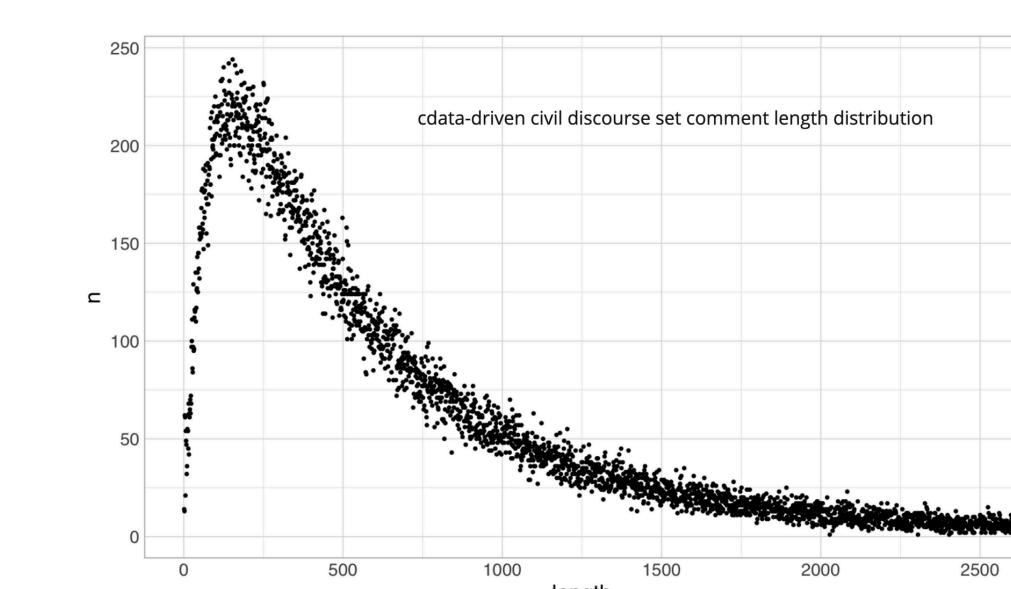
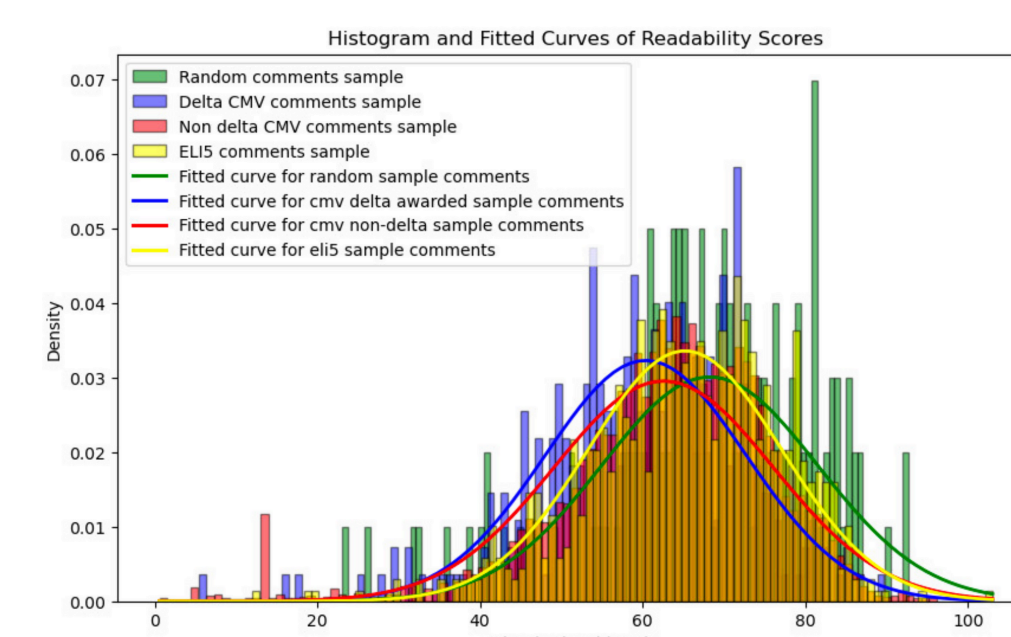
Quality of the data-driven civil set



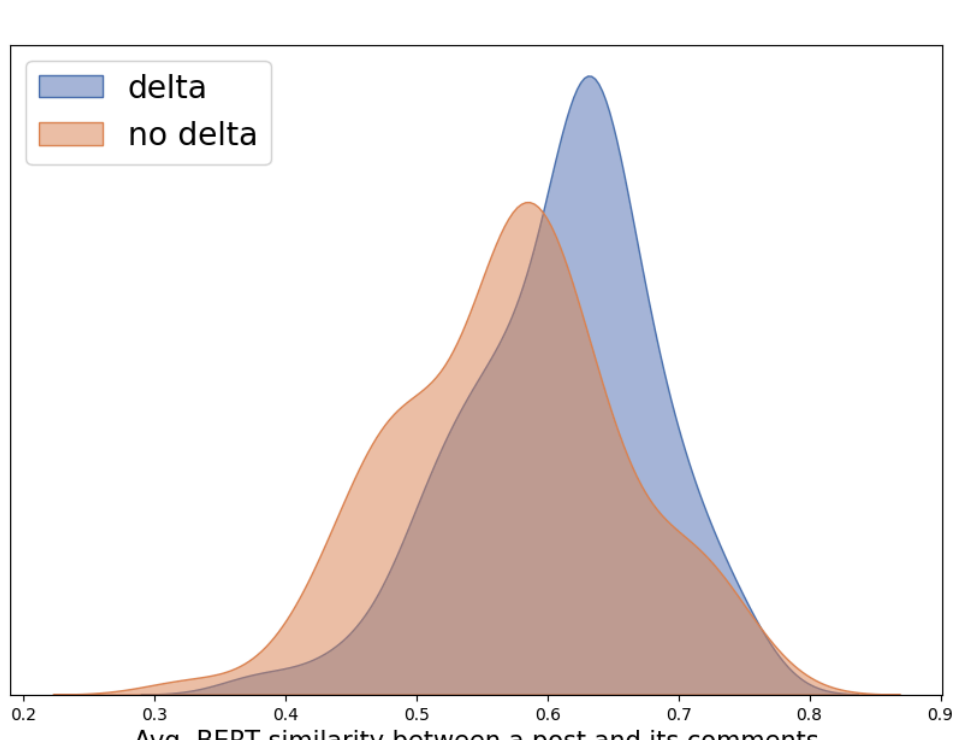
93% of comments in the data-driven civil discourse set human-evaluated as civil

► Should be good for training models, if structural/linguistic features differentiate the subset from other data.

Features differ between the civil set and other data



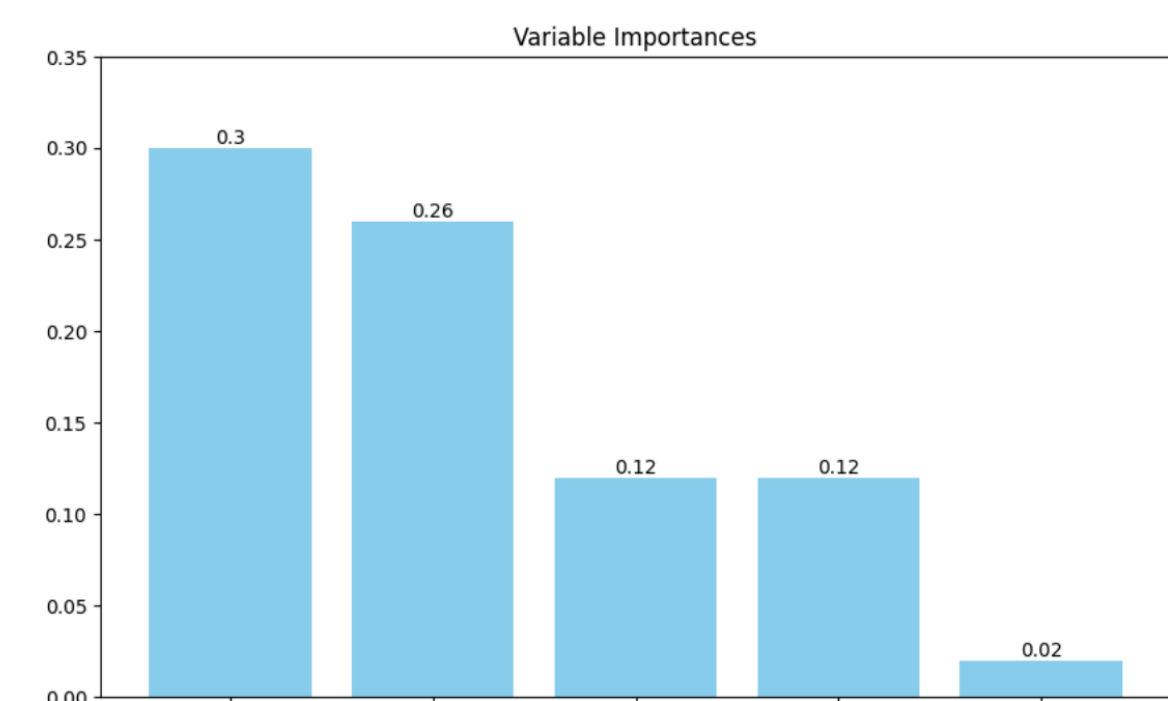
- Comments (esp. delta-awarded) on CMV as compared to random comments tend to be longer, linguistically more complex, and contain more quotes (17% of comments vs. 2%).
- Users who have been awarded many (>150) deltas tend to communicate in a structurally similar way on other subreddits.
- Threads with deltas are semantically more uniform.



A classifier can separate the sets

66%

Precision and recall



A random forest classifier was trained using a small subsample of the civil set and features and validated with the annotated data.

Interventions as civility

Interventions steer conversations to address perceived norm violations. We manually classified instances of *stop arguing*, a common intervention phrase (Tanskanen, 2021).

Genuine interventions

- “Hey let's just *stop arguing* because we agree.”
- “I'll *stop arguing* this point, because I don't think I can change your view...”

Not interventions

- “A friend of mine posted on FB how white people need to *stop arguing* ...”
- “If you're willing to admit that, I'll *stop arguing*.”

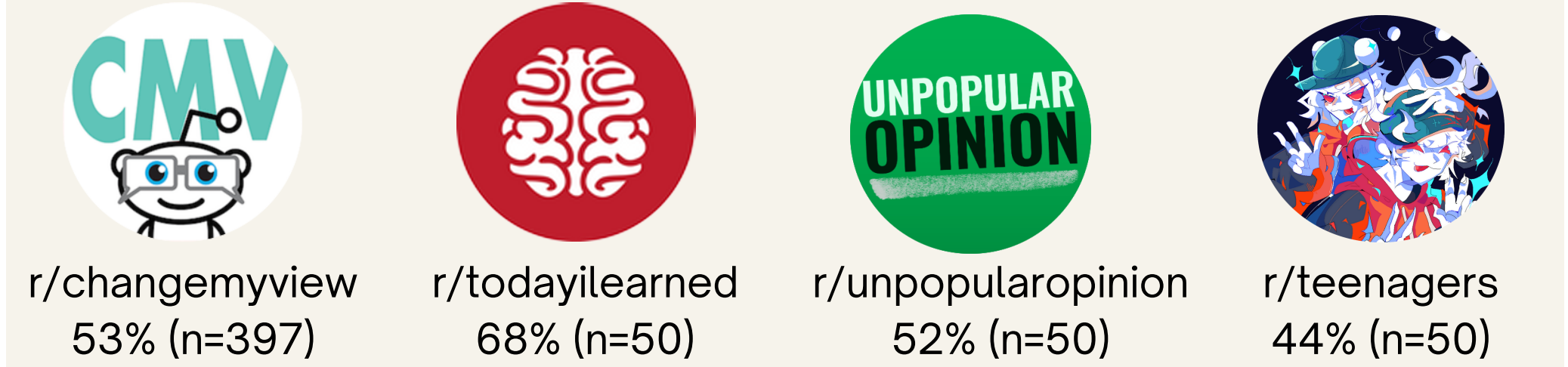
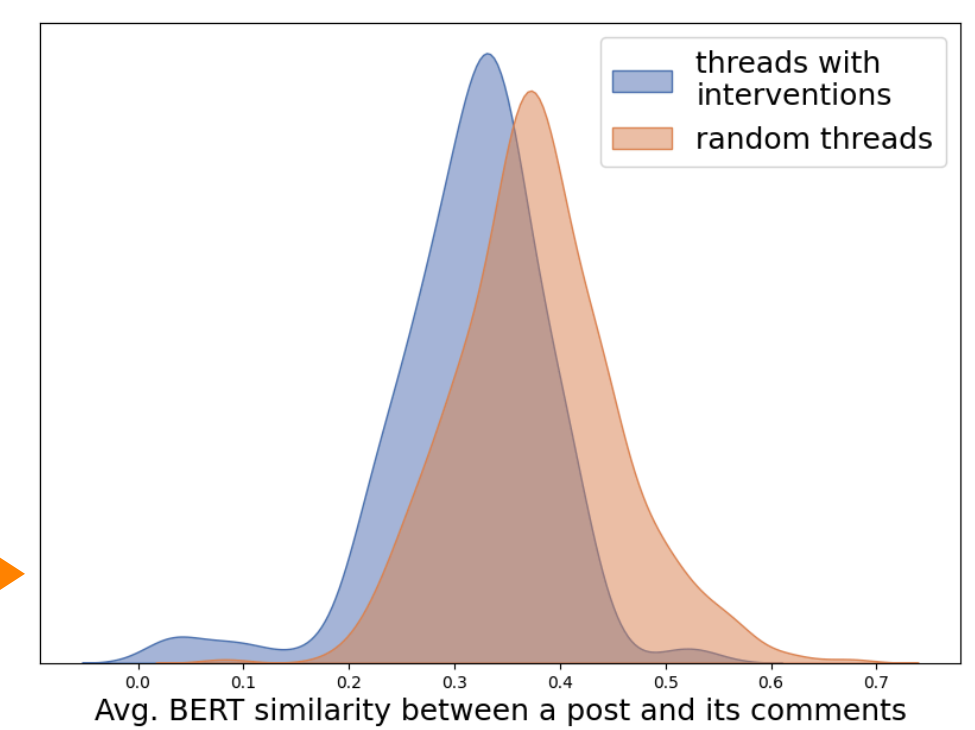
Ambiguous interventions

- “Please *stop arguing* with me, you're unintelligent.”
- “Now *stop arguing* something you know nothing about and do the research yourself if you really care.”

- Interventions uphold r/CMV rules: they stop arguments, refocus discussions, and call out fallacious arguments.
- Civility features predict genuine interventions, but many interventions also contain incivility, thus remaining ambiguous.

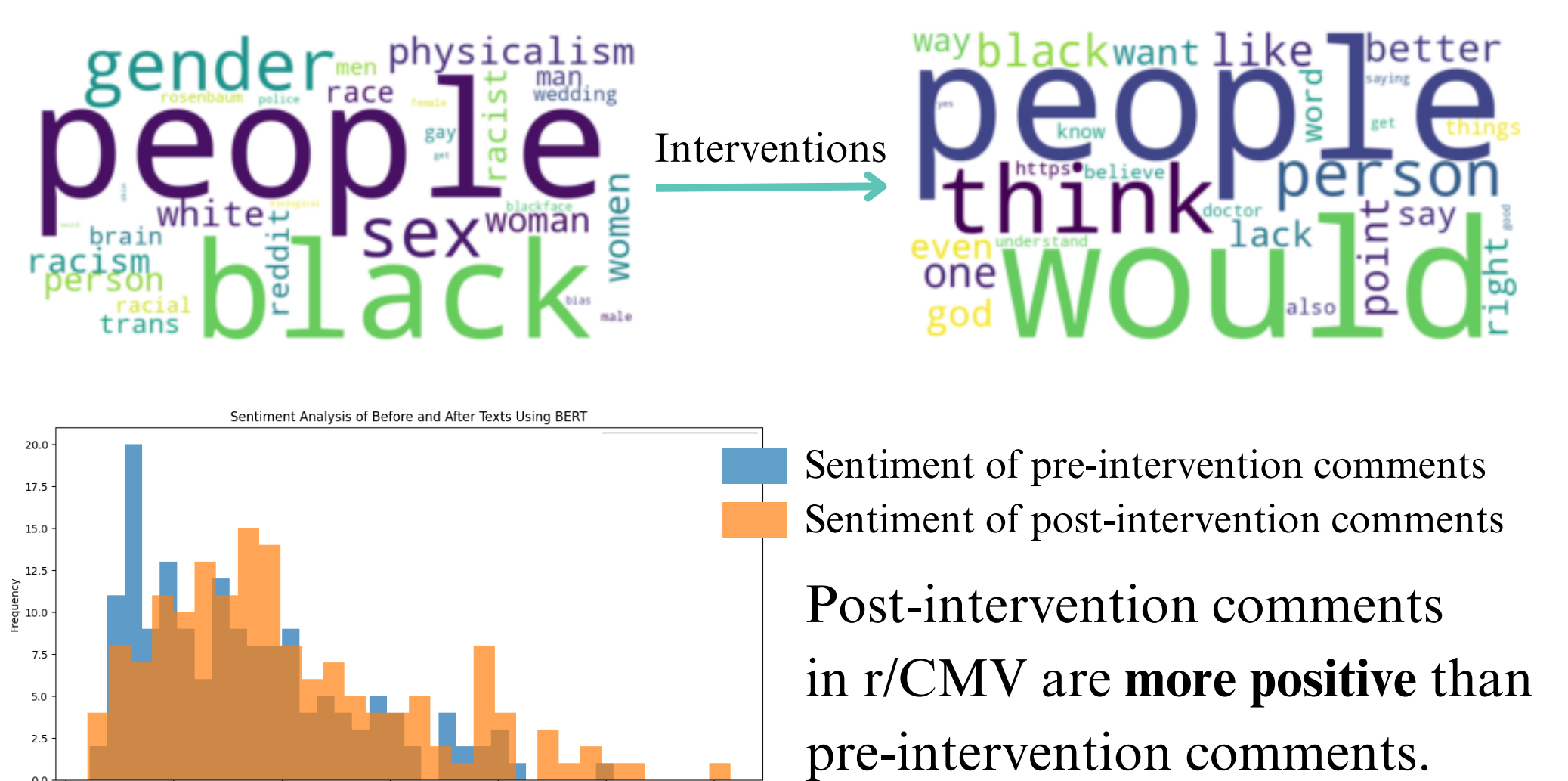
Where do interventions usually occur?

- Semantically divergent threads likely face intervention.
- When discussion is off-topic or goes to extreme viewpoint.
- When discussion is complex or unproductive.
- How often *stop arguing* is a genuine intervention differs by subreddit.



What happens after a genuine intervention?

Topic modeling reveals that post-intervention comments encompass **broader themes**, suggesting that interventions may deter off-topic and extreme viewpoints.



References

- Bonotti, M., & Zech, S. T. (2021). *Recovering civility during COVID-19*. Palgrave Macmillan.
- Hauser, G. (1999). *Vernacular voices: The rhetoric of publics and public spheres*. University of South Carolina Press.
- Kingwell, M. (1995). *A civil tongue: Justice, dialogue and the politics of pluralism*. Pennsylvania State University Press.
- Massaro, T. M., & Stryker, R. (2012). Freedom of speech, liberal democracy, and emerging evidence on civility and effective democratic engagement. *Arizona Law Review*, 54(375).
- Arizona Legal Studies Discussion Paper No. 12-12.
- Papacharissi, Z. (2004). Democracy online: civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society*, 6(2), 259-283.
- Tanskanen, S.-K. (2021). “Stop arguing”: Interventions as metapragmatic acts in discussion forum interaction. In M. Johansson, S.-K. Tanskanen & J. Chovanec (eds.), *Analyzing Digital Discourses: Between Convergence and Controversy*. Palgrave Macmillan, 219-244.