Supplementary materials B3:

Codebook for identifying deliberative norms in topic model generated from Reddit

moderators interview responses.

**Author note:**

This document contains supplementary materials for the article titled:

"Norm repair as an extension to the deliberative dialogue model"

Introduction

This codebook provides instructions for mapping the six norms of online dialogue quality commonly used in deliberative theory (Friess & Eilders, 2015; Goddard & Gillespie, 2023) onto the results of a topic model generated from Reddit moderator responses. The topic model was generated using the BERTopic (Grootendorst, 2022) approach, which clusters similar documents together using sentence embeddings. Our model produced 26 topics (see Figure 1), where each topic is represented by a set of documents, a set of words that indicate those documents, and a title generated from said words and documents from OpenAI's GPT-3.5-turbo generative large language model (OpenAI, 2023; see also Brown et al., 2020). The research question for this study is *How is the deliberative model of online dialogue quality represented in the topic model results?*

Figure 1: Topic model hierarchical clustering of topics.
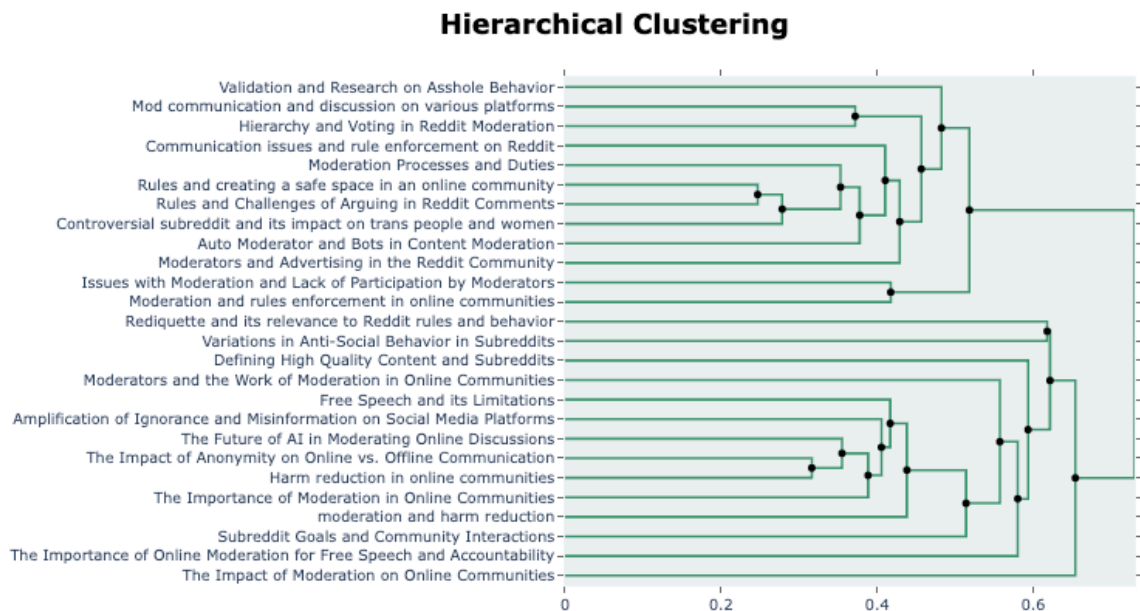
Coding instructions

For coding the variables, we use an Excel spreadsheet with four informative columns providing information about the topic, and six columns referring to the deliberative norms of

dialogue quality. The four informative columns include one with the topic name, one with the words representing a topic, and two columns with a representative document identified by the topic model. The six columns for the coder to score refer to the six norms of online dialogue quality found in the deliberation literature: rationality, civility, inclusiveness, equality, constructiveness, and common good reference (Friess & Eilders, 2015, pp. 328–331). Table 2 provides definitions of each of the norms of deliberative quality.

Table 2: Key aspects of the deliberative norms of online dialogue quality (adapted from Friess & Eilders, 2015, pp. 328–331).

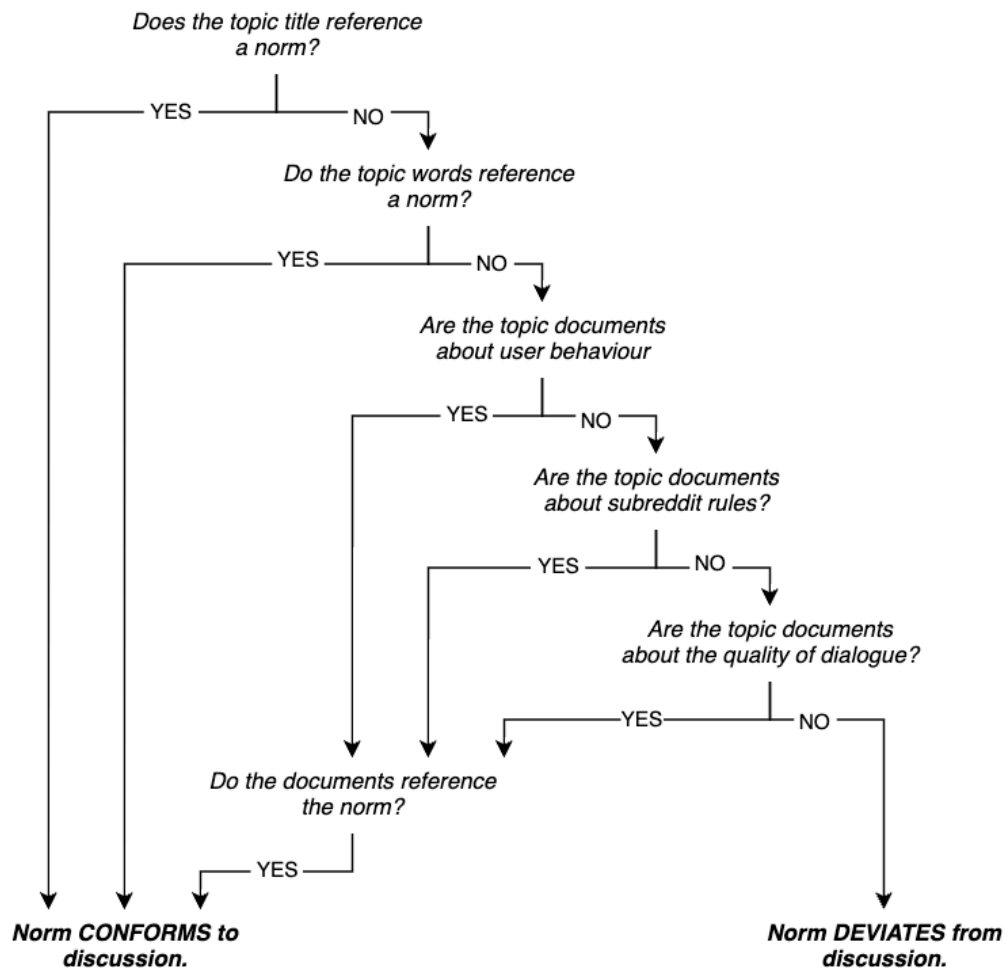| Norm	Description
Rationality	Rationality most commonly refers to the use of "argumentation" by participants in dialogue (Friess & Eilders, 2015, pp. 328–329). However, rationality also refers to participants being *on-topic* during their communications (Stromer-Galley, 2007; Trénel, 2004), being *coherent* in their writing (Graham & Witschge, 2003) and participants justifying their points with facts (Ryfe, 2002). Rationality conforms with a topic if the latter makes any direct reference to participants being on-topic or off-topic, being coherent or incoherent, or any use of argumentation strategies (e.g., justifying their points).
Interactivity	Interactivity refers to the degree of interaction between participants (Friess & Eilders, 2015, pp. 329–330). There are two levels to interactivity (Trénel, 2004). First, "formal interactivity", which refers to whether participants are replying to each other and participating in dialogue (e.g., commenting on someone's post), and, second, "substantial interactivity, which refers to whether participants are *engaging* with each other's points (Friess & Eilders, 2015, p. 329). Engagement can range from showing empathizing with someone, repeating or rephrasing others' perspectives, or engaging in a genuine discussion.
Equality	Equality refers to the Habermasian (1996) notion that anybody who is able or free to participate in the dialogue should have the opportunity to do so (Friess & Eilders, 2015, p. 330). Equality can either refer to discussions of diversity and inclusion of different identity groups (Monnoyer–Smith & Wojcik, 2012), including issues of race, age, gender, education levels, or the degree to which certain participants dominate the discourse.
Civility	Civility refers to the degree of respect (or lack thereof) with which participants treat each other (Friess & Eilders, 2015, p. 330). Civility can vary from indicators of toxic communication (e.g., insults, threats of violence, etc.) to how much participants engage politely and respectfully with each other. Civility can also refer to the degree of empathy and genuineness with which participants engage with each other.

Constructiveness	Constructiveness refers to the degree to which participants are engaging with the goal of reaching some shared outcome, such as agreement or building common ground (Friess & Eilders, 2015, p. 331). Generally, constructiveness is about building consensus between participants (Habermas, 1996). Constructiveness can involve participants' proposing (Trénel, 2004) or discussions (Black et al., 2011) of solutions to a problem in society. It can also be about mobilizing action or using questions to draw out different aspects of a problem (Monnoyer–Smith & Wojcik, 2012).
Common good reference	Common good reference refers to whether participants are in dialogue about the common good (Friess & Eilders, 2015, pp. 330–331). This can include bringing people from different backgrounds and viewpoints together to find common ground and build a better world together (Trénel, 2004). A common good reference is any justification that refers to making things better for a wider group or audience.

Each topic is either coded for whether a norm conforms or deviates from moderators' discussions. The inference of conformity depends on whether the *topic title, topic words, or topic documents* make a reference to the norm or its description. Figure 1 provides a decision tree for when we coded a topic as conforming or deviating. We were only interested in moderators' perspective about *online* dialogue, which refers to the conversations happening on the subreddit or any other public social-media platform (see Goddard & Gillespie, 2023). This means a norm had to be referred to in the context of user behaviors, the rules of a subreddit, or any general discussion about online dialogue and its quality. This also meant any topic about internal discussions between moderators or discussions between administrators were scored as deviating.

Figure 1: Decision tree for assessing conformity between norms and moderators' discussion.

Does the topic title reference a norm?

YES ——— | ——— NO

Do the topic words reference a norm?

YES ——— | ——— NO

Are the topic documents about user behaviour

YES ——— | ——— NO

Are the topic documents about subreddit rules?

YES ——— | ——— NO

Are the topic documents about the quality of dialogue?

YES ——— | ——— NO

Do the documents reference the norm?

YES

Norm CONFORMS to discussion.

Norm DEVIATES from discussion.

References:

Black, L. W., Welser, H. T., Cosley, D., & DeGroot, J. M. (2011). Self-governance through group discussion in Wikipedia: Measuring deliberation in online groups. *Small Group Research*, *42*(5), 595–634. https://doi.org/10.1177/1046496411406137

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901. https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html

Friess, D., & Eilders, C. (2015). A systematic review of online deliberation research. *Policy & Internet*, *7*(3), 319–339. https://doi.org/10.1002/poi3.95

Goddard, A., & Gillespie, A. (2023). Textual indicators of deliberative dialogue: A systematic review of methods for studying the quality of online dialogues. *Social Science Computer Review*, *41*(6), 2364–2385. https://doi.org/10.1177/08944393231156629

Graham, T., & Witschge, T. (2003). In search of online deliberation: Towards a new method for examining the quality of online discussions. *Communications*, *28*(2), 173–204. https://doi.org/10.1515/comm.2003.012

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure* (arXiv:2203.05794). arXiv. https://doi.org/10.48550/arXiv.2203.05794

Habermas, J. (1996). *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. The MIT Press.

Monnoyer–Smith, L., & Wojcik, S. (2012). Technology and the quality of public

    deliberation: A comparison between on and offline participation. *International*

    Journal of Electronic Governance, *5*(1), 24–49.

    https://doi.org/10.1504/IJEG.2012.047443

OpenAI. (2023). *ChatGPT* (22 August) [Python]. https://chat.openai.com

Ryfe, D. (2002). The practice of deliberative democracy: A study of 16 deliberative

    organizations. *Political Communication*, *19*(3), 359–377.

Stromer-Galley, J. (2007). Measuring deliberation's content: A coding scheme. *Journal of*

    Deliberative Democracy, *3*(1), Article 1. https://doi.org/10.16997/jdd.50

Trénel, M. (2004). *Measuring the deliberativeness of online discussions. Coding scheme 2.2*

    [Unpublished paper].