



**Selected Papers of #AoIR2024:
The 25th Annual Conference of the
Association of Internet Researchers**
Sheffield, UK / 30 Oct - 2 Nov 2024

UNDERSTANDING PERCEPTIONS AND EFFECTS OF ONLINE INTOLERANCE: A FOUR-COUNTRY EXPERIMENTAL STUDY

Patrícia Rossini
University of Glasgow

Cristian Vaccari
University of Edinburgh

Yannis Theocharis
Technical University Munich

Rebekah Tromble
George Washington University

Introduction

Research on online political discourse has long been concerned with the pervasiveness of incivility across various digital arenas, considering it a problematic feature of public speech (Chen et al., 2019). While empirical work across different platforms has suggested that incivility is frequently encountered online, a key limitation of this work is its empirical focus on rude, vulgar, or offensive content, which may not inherently be perceived as incompatible with political discussions (Chen et al., 2019; Muddiman, 2017; Stryker et al., 2016). While these discursive features may be seen as rude (Coe et al., 2014), they are not necessarily harmful to discussion participants (Rossini, 2022). In other words, work on incivility has been primarily concerned with the tone, not the substance, of online discourse. As a result, we still know very little about how discourse that is truly harmful affects both targets and bystanders in online platforms.

To address this gap, we focus on intolerant online discourse—content that is hateful, threatening, discriminatory, or harassing (Rossini, 2022). Investigating the potential effects of exposure to intolerant speech is important for three main reasons: First, it helps understand how perceptions of intolerance may vary based on its severity—for instance, discrimination is typically disregarded as a violation of ‘community standards’, while hateful or threatening speech tends to be moderated. Second, it helps identify

Suggested Citation (APA): Rossini, P., Vaccari, C., Theocharis, Y., Tromble, R. (2024, October). *Understanding Perceptions and Effects of Online Intolerance: A Four-Country Experimental Study*. Paper presented at AoIR2024: The 25th Annual Conference of the Association of Internet Researchers. Sheffield, UK: AoIR. Retrieved from <http://spir.aoir.org>.

those who are most vulnerable—both as *targets* and *bystanders*— to intolerance. Finally, it helps understand how these groups react to intolerant speech—e.g., by hiding, blocking, or reporting the authors, but also by reducing their own activity on social media, such as avoiding discussion topics, posting less frequently, reducing interactions, or even deleting their profile—thus quantifying the harm that intolerant speech causes. Importantly, this approach highlights the detrimental effects of intolerant speech on bystanders, who tend to be overlooked by community standards and moderation practices focused on targets.

Intolerant discourse is a serious threat to online political expression: according to the Pew Research Center (Anderson & Quinn, 2019), 41% of Americans report having experienced online harassment, with the majority of the cases taking place on social media, while 1 in 4 teens in the UK reported having seen hateful messages (Ofcom, 2019). In Germany, 40% of internet users reported exposure to online hate speech (Geschke et al., 2019), while in Brazil the organization Safernet¹ reported in 2022 a substantial rise in the reporting of hateful online crimes, including racism, xenophobia, misogyny, and religious intolerance. Intolerant content can silence marginalized voices and may turn social media platforms into hostile spaces, undermining democratic discourse, and leading people to abandon them as places for discussion. However, limited research has investigated the negative effects of being targeted or exposed to online intolerance.

In this study, we tackle this important research gap and investigate the potential effects of exposure to intolerant speech in four countries where online intolerance is pervasive—Brazil, Germany, the UK, and the US. The four countries in our sample represent democracies with highly active internet users and feature a high level of political animosity towards disadvantaged groups, such as immigrants, LGBTQ+, and other minorities. Notably, far-right parties and politicians have weaponized discourse towards minorities in recent electoral cycles across all these countries. These countries also differ in how they regulate online speech, with Germany and Brazil having more restrictive legislation, and the UK and the US as the least regulated regimes.

Methods

We leverage pre-registered survey experiments on large samples (N = 2,000 per country) constructed to mirror the adult population on key characteristics in these countries to examine perceptions of and reactions to intolerant online discourse. Our experiments manipulate the target (women, LGBT), tone (civil, uncivil), and type (discriminatory, hateful, threatening) of intolerant online discourse, using realistic mock-ups of social media posts. Our key dependent variables include (1) perceptions of how harmful the content is; (2) intent to react to the post (e.g., by reporting, blocking, responding); and (3) support for content moderation practices. Based on existing research investigating both perceptions and effects of online incivility, we can expect reactions to intolerant speech will vary based on personal traits, political attitudes, and experiences with online toxicity, as well as being a member of a 'targeted' identity

¹ <https://new.safernet.org.br/content/denuncias-de-imagens-de-abuso-e-exploracao-sexual-infantil-reportadas-pela-safernet>

group. We also observe whether these effects are consistent across different countries and contexts.

We manipulated different types of intolerance because we expect participants to respond differently to them based on perceived harmfulness. For instance, discrimination often falls outside the scope of community standards and platform moderation rules, while hate speech and violent threats are likely perceived as more harmful (Stryker et al., 2016). As such, we expect participants to be more sensitive to hateful and threatening speech, with discrimination being perceived as less problematic than the other types (H1). We also expect the tone of the message to have an impact, with intolerant discourse that is uncivil being perceived more harshly than intolerant discourse presented in a civil tone (H2). These differences are also likely to shape intentions to react to posts in the same direction—e.g., responding, blocking, or reporting, as well as reducing engagement with political discussions (H3). Finally, we expect participants exposed to the treatments to report higher support for content moderation practices in the uncivil condition (H4a), as well as in the more harmful types of intolerance compared to discrimination (H4b). Considering the position of participants as targets or bystanders, we expect participants' identification as a member of the targeted groups (women, LGBT+) to be more affected by our treatments, regardless of the tone of the treatment (H5).

References

- Anderson, M., & Quinn, D. (2019, August 8). 46% of social media users worn out by political posts, discussions. *Pew Research Center*.
<https://www.pewresearch.org/fact-tank/2019/08/08/46-of-u-s-social-media-users-say-they-are-worn-out-by-political-posts-and-discussions/>
- Chen, G. M., Muddiman, A., Wilner, T., Pariser, E., & Stroud, N. J. (2019). We Should Not Get Rid of Incivility Online. *Social Media + Society*, 5(3), 2056305119862641. <https://doi.org/10/gghcnh>
- Coe, K., Kenski, K., & Rains, S. A. (2014). Online and Uncivil? Patterns and Determinants of Incivility in Newspaper Website Comments. *Journal of Communication*, 64(4), 658–679. <https://doi.org/10/f6dxrx>
- Geschke, D., Klößen, A., Quent, M., & Richter, C. (2019). *#HASS IM NETZ: Der schleichende angriff auf unsere demokratie*.
- Muddiman, A. (2017). Personal and Public Levels of Political Incivility. *International Journal of Communication*, 11(0), 21.
- Ofcom. (2019). *Children and parents: Media use and attitudes report*.
- Rossini, P. (2022). Beyond Incivility: Understanding Patterns of Uncivil and Intolerant Discourse in Online Political Talk. *Communication Research*, 49(3), 399–425. <https://doi.org/10/gkfj98>

Stryker, R., Conway, B. A., & Danielson, J. T. (2016). What is political incivility?
Communication Monographs, 83(4), 535–556. <https://doi.org/10/dz7s>