How Hostile Tone Shapes Science Communication

No Author Given

No Institute Given

Abstract. Scientists face growing online hostility, yet little is known about how hostile science content behaves on social media platforms or how exposure to it affects audiences. This paper presents a multi-method study combining large-scale observational Twitter data ($N = 24{,}135{,}095$ tweets) with two pre-registered online experiments ($N = 88$; $N = 279$) and a large-scale trust study ($N = 2{,}398$). Our findings reveal that hostile tone suppresses perceived accuracy and sharing intentions at the individual level. On Twitter, however, hostile science content circulates more widely than neutral content. This divergence reveals how method choice shapes inference, and we argue that triangulating across levels of analysis is essential for valid computational social science research on online hostility.

Keywords: science communication · online hostility · social media analytics · multi-method design · computational social science.

1 Motivation

Scientists face significant online hostility: roughly 30% of academic staff report harassment within a six-month period (Oksanen et al., 2022), and 81% of COVID-19 researchers who engaged with media received personal attacks (Nogrady, 2021). Computational social science is well positioned to study this phenomenon at scale, but the field faces a recurring methodological challenge: large-scale behavioral data and individual-level experiments do not measure the same constructs, and treating them as interchangeable can produce contradictory or misleading conclusions. This paper uses hostile science communication as a substantive case to demonstrate this challenge and to propose multi-method triangulation as a principled response.

2 Multi-Method Framework

2.1 Platform level

We analyzed 24,135,095 tweets linking to academic articles (2011–2019), drawn from Zhang & Horvát (2025). Tweets were classified using Google's Perspective API; 97,125 (0.40%) were offensive. We further classified the tweets specifically targeting scientists with a combination of key-word-based search, GPT-4o-assisted labeling, and human evaluation. The final number of 1,534 hostile

tweets targeting scientists were only included if two human annotators agreed that they were targeted towards scientists. Science-targeting offensive tweets received substantially more engagement than non-offensive science tweets: mean 2.73 retweets and 6.38 likes versus 1.44 and 2.28 ($p < .001$; see Appendix: Figure 1). This pattern is consistent with moral-emotional diffusion accounts (Brady et al., 2017), which predict that anger-inducing content is amplified algorithmically and through reactive engagement.

2.2 Individual level

To establish how individual users actually evaluate hostile content, we ran two pre-registered experiments in which participants read science headlines manipulated on tone (hostile vs. neutral). Study 1a ($N = 88$) found that hostile tone reduced perceived accuracy by 1.15 points and sharing intentions by 0.68 points (both $p < .001$; see Appendix: Figures 2 3). Study 1b ($N = 279$) replicated both effects and additionally crossed tone with political endorsement; aligned source endorsement did not attenuate the hostility penalty.

2.3 Institutional level

A between-subjects design ($N = 2{,}398$) exposed participants to either neutral headlines or a mix of hostile and neutral headlines, measuring general trust in science before and after using a validated short version of the Trust in Science and Scientists Scale scale (Nadelson et al., 2014). Trust scores were stable across conditions (pre $M = 5.30$, post $M = 5.29$; $t = -0.62$, $p = .54$), with no significant difference between the hostile and neutral arms ($p = .58$).

3 Conclusion

The three data sources, taken individually, tell incomplete or contradictory stories. The Twitter data suggests hostile content is rewarded by the ecosystem: offensive science tweets received nearly twice the retweets and three times the likes of non-offensive ones. The experimental data points in the opposite direction: accuracy judgments dropped by over a point and sharing intentions followed.

These are not contradictions to be resolved by choosing the better method; they reflect genuinely distinct phenomena operating at different scales. Platform-level engagement aggregates across algorithmic dynamics, network structure, and heterogeneous motivations. Individual-level experiments isolate epistemic evaluation under controlled single-exposure conditions, which aggregate counts cannot distinguish. The trust study targets a construct that moves more slowly than item-level evaluation—if at all. Taken together, the three levels show that the same hostile content can spread widely on a platform, be judged as less accurate by individuals, and leave institutional trust intact.

Bibliography

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *PNAS*, *114*(28), 7313–7318.

Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology*, *81*(1), 146–159.

Nadelson, L., Jorcyk, C., Yang, D., Jarratt Smith, M., Matson, S., Cornell, K., & Husting, V. (2014). I just don't trust them: The development and validation of an assessment instrument to measure trust in science and scientists. *School Science and Mathematics*, *114*(2), 76–86.

Nogrady, B. (2021). 'I hope you die': how the COVID pandemic unleashed attacks on scientists. *Nature* News Feature. https://www.nature.com/articles/d41586-021-02741-x

Oksanen, A., Celuch, M., Latikka, R., Oksa, R., & Savela, N. (2022). Hate and harassment in academia: the rising concern of the online environment. *Higher Education*, *84*(3), 541–567.

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39–50.

Steger, D., & Schütz, A. (2025). Assessing trust in science: Development and validation of a short scale for adolescents and adults. *Measurement Instruments for the Social Sciences*, *7*, e16225.

Zhang, Y., & Horvát, E.-Á. (2025). The unwanted dissemination of science: The usage of academic articles as ammunition in contested discursive arenas on Twitter. *Proceedings of the ACM on Human-Computer Interaction*, *9*(7), 1–24.

4 Appendix

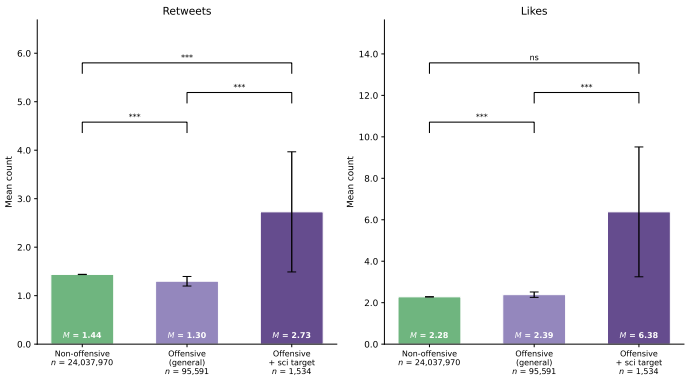

Fig. 1. Mean engagement by tweet group (± SE). Asterisks denote significance levels from negative binomial regression (*** $p < .001$).

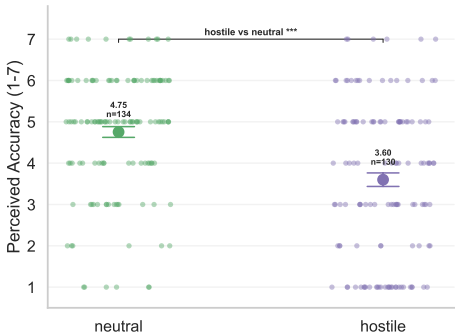

Fig. 2. Perceived accuracy by tone (Study 1a). Differences estimated via t-test; *** $p < .001$.

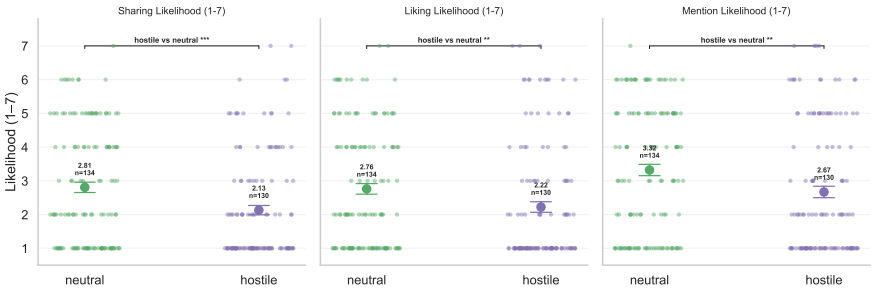

Fig. 3. Engagement intentions by tone (Study 1a). Differences estimated via t-test; *** $p < .001$.