Overuse of moral language dampens content engagement on social media

Keywords: Morality, Social Media, Moral Engagement, Text Analysis, Politics.

Extended Abstract

Online social media platforms are a vital arena in which socio-political perspectives are put forth, debated, and propagated [1]. Researchers dispute what drives the spread of certain messages over others on these platforms. Using data from a single platform (i.e., Twitter), one group suggests that moral-emotional language is a key driver of online contagion [2], but others have questioned this conclusion [3]. We substantially advance this research using a broader range of topics and social media platforms, including two mainstream social media platforms (Twitter and Reddit) and one unmoderated platform popular among right-wing extremists (8Chan).

We collected data on 2,141,933 online posts. On Twitter, we analyze the texts from tweets (953,179) and use the number of retweets to index engagement. On Reddit (1,067,832 posts) and 8 Chan (120,922 posts), we analyze opening texts from threads—where people have the chance to write replies—and we use the number of replies (including the number of replies to replies in Reddit) that each thread generates to measure engagement. In all platforms we count the number of moral words using the Moral Foundation Dictionary [4] as well as the number of total words. After controlling for confounders at the level of posts such as text length, URLs, media content, number of followers, our findings demonstrate two countervailing mechanisms that predict engagement with online posts. We confirm the idea that infusing content with moral language increases engagement (Figure 1A-D). However, in contrast to prior work, we find that this effect is sub-linear rather than exponential and find no evidence that it is specifically driven by moral-emotional (vs. moral) language. We also identify a novel negative effect of morality on engagement (Figure 1E-H). Specifically, we find that a higher ratio of moral to non-moral words in a message markedly decreases engagement: Holding the number of moral words present in a given online post constant, a higher proportion of moral to non-moral words dampens online diffusion across socio-political issues, a phenomenon we call "moral saturation."

We run a set of robustness checks. First, we find evidence of divergent validity. Most of the coefficients are not significant when applying the same model to neutral words rather than moral words, highlighting the specificity of our model to moral content per se (i.e., neutral words do not exhibit the patterns described in Figure 1). For the Reddit data, we also examine an alternative measure of engagement, the posts' score—the difference between the number of times a post was upvoted versus downvoted. Results suggest that the moral saturation pattern holds even for an alternative measure of online engagement. Finally, we address the concern that the Moral Foundations Dictionary does not weight moral words by their likelihood of being used in clearly moral contexts. To address this issue, we replicate our results using a distributed dictionary representation [5], which relies on semantic similarity using word embeddings. Thus, we unveil a previously unknown pattern of moral saturation that shows up in all analyzed online platforms. Mathematically, moral saturation leads to an optimal level of moralization that maximizes the online content engagement.

These results provide insights on how people engage in moral content in the digital era and generalize previous research across a variety of online platforms and themes. Collectively, our work clarifies the complex role of morality in driving engagement of online content: infusing messages with moral language increases their spread to a point but relying excessively on morality may backfire.

References

[1] Del Vicario, Michela, et al. "The spreading of misinformation online." *Proceedings of the National Academy of Sciences* 113.3 (2016): 554-559.
[2] Brady, William J., et al. "Emotion shapes the diffusion of moralized content in social networks." *Proceedings of the National Academy of Sciences* 114.28 (2017): 7313-7318.
[3] Burton, Jason W., Nicole Cruz, and Ulrike Hahn. "Reconsidering evidence of moral contagion in online social networks." *Nature Human Behaviour* 5.12 (2021): 1629-1635
[4] Graham, Jesse, Jonathan Haidt, and Brian A. Nosek. "Liberals and conservatives rely on different sets of moral foundations." *Journal of personality and social psychology* 96.5 (2009): 1029.
[5] Garten, Justin, et al. "Dictionaries and distributions: Combining expert knowledge and large scale textual data content analysis." *Behavior research methods* 50.1 (2018): 344-361.

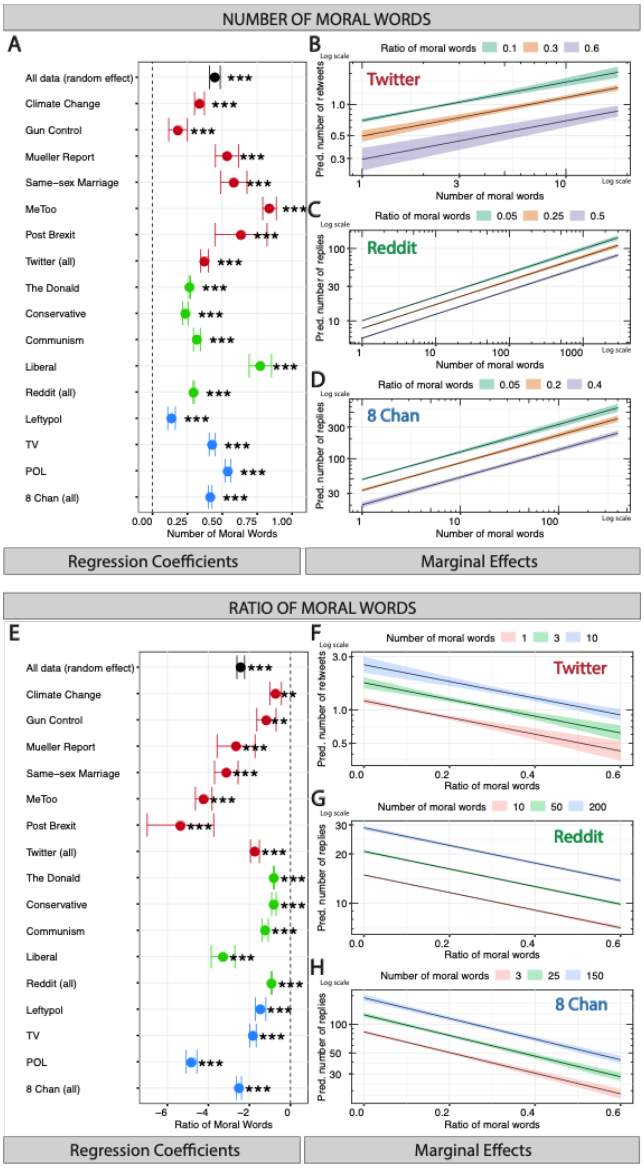

Figure 1. Moralization leads to more engagement (top panel) but the overuse of moral language dampens content engagement (bottom panel).