

Norm Psychology in the Digital Age: How Social Media Shapes the Cultural Evolution of Normativity

William J. Brady¹  and M. J. Crockett² 

¹Kellogg School of Management, Northwestern University, and ²Department of Psychology and University Center for Human Values, Princeton University

How do people think about what is appropriate, allowed, required, or forbidden? Heyes’s “cognitive gadgets” account of norm psychology proposes that cultural selection affects not just the content of norms but also how people think and feel about norms in general. Contra nativist views that paint human normativity as genetically inherited and evolving at a glacial pace, Heyes’s model suggests that people alive today have substantially more control over how future generations think about norms and their enforcement.

Here, we explore one important implication of Heyes’s account: New technologies, such as social media, can spur rapid cultural evolution of normativity (Acerbi, 2019). Commentary plays a key role in the cultural transmission of norms by defining the terms of compliance and enforcement, and it is increasingly taking place online, with more than 4 billion social media users worldwide (Dixon, 2022). Building on recent studies of online normative discourse, we speculate how three features of social media shape the cultural evolution of normativity: hidden incentives, multiple audiences, and speed of transmission.

By incentivizing particular types of commentary over others in ways that are not always transparent to users, social media can teach people new ways to think and feel about norms. For example, preliminary evidence suggests that social media incentivizes moral outrage. Because social media platforms profit more by keeping users online, their algorithms promote content that is most likely to be shared. Users are especially likely to share commentary containing moralized and emotional language (e.g., moral outrage) because it attracts attention and highlights group identity (Brady, Crockett, & Van Bavel, 2020; Brady et al., 2017, 2019; Brady, Gantman, & Van Bavel, 2020; Brady & Van Bavel, 2021; Crockett, 2017), and algorithms amplify this tendency (Brady et al., in press; Chakradhar, 2021; Chowdhury, 2021; Huszár et al., 2022). The widespread sharing of moral

outrage has two consequences: It teaches users that outrage expressions are normative, and it directly reinforces those outrage expressions. Thus, users learn over time to express more outrage (Brady et al., 2021), which further inflates beliefs that outrage is normative. Inflated prior beliefs about the prevalence of outrage may then lower the threshold for perceiving it, causing individuals to believe others are more outraged than they actually are (Brady et al., 2023).

We speculate that these algorithmic social-learning processes could extend beyond outrage expressions to affect other types of normative thinking. For example, simplified moral narratives with a clear villain and victim may generate more engagement than more nuanced accounts. Because social media algorithms gamify engagement, individuals may learn over time to communicate and perhaps even to cognitively represent moral situations in ways that essentialize moral character, which in turn increases harsh punitive attitudes (Heiphetz, 2020). Moreover, there is evidence that social media algorithms disproportionately incentivize and spread right-wing over left-wing political commentary in the United States (Huszár et al., 2022). Such incentives, over time, could promote not just the content of right-wing political views but also right-wing styles of normative thinking, such as moral parochialism (Bai et al., 2021; Kahane et al., 2018; Waytz et al., 2019). Furthermore, because social media algorithms disproportionately incentivize criticism directed toward out-groups (Rathje et al., 2021), this could increase the impact of social identity on normative thinking, making people quicker to blame and assume the worst intentions of out-group members. Such a dynamic would

Corresponding Author:

M. J. Crockett, Department of Psychology and University Center for Human Values, Princeton University
 Email: mj.crockett@princeton.edu

reinforce the same negative out-group metaperceptions that exacerbate conflict (Moore-Berg et al., 2020; Ruggeri et al., 2021).

A second consequential feature of social media is the way it flattens multiple audiences into one, a phenomenon known as “context collapse” (Marwick & Boyd, 2011). Both theory and evidence underscore the role of imagined audiences in directing people’s normative behaviors: Compliance, enforcement, and commentary are all sensitive to reputational concerns even when no one is watching (Jordan & Rand, 2020; Kim & Crockett, 2022; Smith, 1822). Context collapse on social media substantially complicates reputation management because commentary meant for one particular audience might end up being seen by a very different audience, perhaps even one that the commenter did not even know existed. To navigate this, commenters must hold very different conceptions of their imagined audience in mind when engaging in online commentary compared with offline conversations in which the audience is usually fully known. There is evidence that at least some social media users develop strategies to deal with context collapse, such as self-censorship (Marwick & Boyd, 2011). As these strategies are shared, they may spur the cultural evolution of increasingly sophisticated reputation-management strategies in commentary that are optimized for digital spaces.

On social media, information spreads through social networks at an unprecedented speed, which in turn can substantially increase the rate of social learning. Users can observe thousands of social reactions to normative commentary every time they log on to social media and receive social feedback on their own commentary from thousands of people within minutes. The rapid evolution of norms on social media may not just accelerate the adoption of new types of normative thinking but also affect people’s beliefs about the functions and legitimacy of norms themselves. For example, viral outrage “pile-ons” increase beliefs that outrage is normative but also increase sympathy for the offender because the outrage is seen as disproportionate to the offense (Sawaoka & Monin, 2018). Observing how certain types of viral normative commentary can quickly grow a user’s audience may promote cynical beliefs that people primarily express particular moral views *in order to* bolster their reputation regardless of what they actually believe (Kraft-Todd et al., in press; Tosi & Warmke, 2020). More broadly, observing norms evolve online at a bewildering pace, against a historical backdrop of much slower normative change, might encourage a sort of “normative nihilism”: If it is too difficult to keep track of what is normative and what is not, some people might conclude that modern morality has lost any meaningful function and retreat to more traditional norms that they perceive to be more legitimate.

The biologist E. O. Wilson (<https://www.harvardmagazine.com/breaking-news/james-watson-edward-o-wilson-intellectual-entente>) famously ascribed the problems of humanity to “Paleolithic emotions, medieval institutions and godlike technology.” Nativist accounts of normativity suggest people’s present crises stem from an inevitable mismatch between their ancient moral emotions and modern technologies and propose solutions that work around people’s misfiring moral intuitions rather than seeking to change them. The cognitive-gadgets account is more optimistic: It suggests that people’s norm psychology can change on a much more rapid timescale and that people have much more agency and control over how that change proceeds. Here, we discussed how features of online environments can shape the cultural evolution of normativity and suggested that social media design is not currently optimized for cultural selection of gadgets that promote cooperation. But this consequence is not inevitable. We hope that increasing attention to how normativity is shaped by culture—in addition to nature and nurture—can pave the way for designing technologies that accelerate the evolution of a norm psychology better adapted for global cooperation.

Transparency

Action Editor: Daniel Kelly

Editor: Interim Editorial Panel

Declaration of Conflicting Interests


The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

M. J. Crockett was supported by grants from the John Templeton Foundation (61495 and 62816) and Democracy Fund (R-201809-03031).

ORCID iDs

William J. Brady  <https://orcid.org/0000-0001-6075-5446>

M. J. Crockett  <https://orcid.org/0000-0001-8800-410X>

References

- Acerbi, A. (2019). *Cultural evolution in the digital age*. Oxford University Press.
- Bai, X., Gauri, V., & Fiske, S. T. (2021). Cosmopolitan morality trades off in-group for the world, separating benefits and protection. *Proceedings of the National Academy of Sciences, USA, 118*(40), Article e2100991118. <https://doi.org/10.1073/pnas.2100991118>
- Brady, W. J., Crockett, M. J., & Van Bavel, J. J. (2020). The MAD model of moral contagion: The role of motivation, attention, and design in the spread of moralized content online. *Perspectives on Psychological Science, 15*(4), 978–1010. <https://doi.org/10.1177/1745691620917336>
- Brady, W. J., Gantman, A. P., & Van Bavel, J. J. (2020). Attentional capture helps explain why moral and emotional content go viral. *Journal of Experimental Psychology:*

- General*, 149(4), 746–756. <https://doi.org/10.1037/xge0000673>
- Brady, W. J., McLoughlin, K., Doan, T. N., & Crockett, M. (2021). How social learning amplifies moral outrage expression in online social networks. *Science Advances*, 7(33), Article eabe5641. <https://doi.org/10.1126/sciadv.abe5641>
- Brady, W. J., Jackson, J. C., Lindström, B., & Crockett, M. J. (in press). Algorithm-mediated social learning in online social networks. *Trends in Cognitive Sciences*.
- Brady, W. J., McLoughlin, K., Torres, M., Luo, K., Gendron, M., & Crockett, M. (2023). Overperception of moral outrage in online social networks inflates beliefs about intergroup hostility. *Nature Human Behaviour*, 7(6), 917–927.
- Brady, W. J., & Van Bavel, J. J. (2021). *Social identity shapes antecedents and functional outcomes of moral emotion expression in online networks*. OSF Preprints. <https://doi.org/10.31219/osf.io/dgt6u>
- Brady, W. J., Wills, J. A., Burkart, D., Jost, J. T., & Van Bavel, J. J. (2019). An ideological asymmetry in the diffusion of moralized content on social media among political leaders. *Journal of Experimental Psychology: General*, 148(10), 1802–1813. <https://doi.org/10.1037/xge0000532>
- 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. *Proceedings of the National Academy of Sciences, USA*, 114(28), 7313–7318. <https://doi.org/10.1073/pnas.1618923114>
- Chakradhar, S. (2021). *More internal documents show how Facebook's algorithm prioritized anger and posts that triggered it*. Nieman Lab. <https://www.niemanlab.org/2021/10/more-internal-documents-show-how-facebooks-algorithm-prioritized-anger-and-posts-that-triggered-it/>
- Chowdhury, R. (2021). *Examining algorithmic amplification of political content on Twitter*. https://blog.twitter.com/en_us/topics/company/2021/rml-politicalcontent
- Crockett, M. J. (2017). Moral outrage in the digital age. *Nature Human Behaviour*, 1(11), Article 11. <https://doi.org/10.1038/s41562-017-0213-3>
- Dixon, S. (2022, September 1). *Number of worldwide social network users 2027*. Statista. <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>
- Heiphetz, L. (2020). The development and consequences of moral essentialism. *Advances in Child Development and Behavior*, 59, 165–194.
- Huszár, F., Ktena, S. I., O'Brien, C., Belli, L., Schlaikjer, A., & Hardt, M. (2022). Algorithmic amplification of politics on Twitter. *Proceedings of the National Academy of Sciences, USA*, 119(1), Article e2025334119. <https://doi.org/10.1073/pnas.2025334119>
- Jordan, J. J., & Rand, D. G. (2020). Signaling when no one is watching: A reputation heuristics account of outrage and punishment in one-shot anonymous interactions. *Journal of Personality and Social Psychology*, 118(1), 57–88.
- Kahane, G., Everett, J. A., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2018). Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychological Review*, 125(2), 131–164.
- Kim, J., & Crockett, M. J. (2022). Narrating the “what” and “why” of our moral actions. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 44. <https://escholarship.org/uc/item/9xk406n>
- Kraft-Todd, G., Kleiman-Weiner, M., & Young, L. (in press). *Virtue discounting: Observability reduces moral actors' perceived virtue*. Open Mind.
- Marwick, A. E., & Boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1), 114–133.
- Moore-Berg, S. L., Ankori-Karlinisky, L. O., Hameiri, B., & Bruneau, E. (2020). Exaggerated meta-perceptions predict intergroup hostility between American political partisans. *Proceedings of the National Academy of Sciences, USA*, 117(26), 14864–14872.
- Rathje, S., Van Bavel, J. J., & Van Der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences, USA*, 118(26), Article e2024292118. <https://doi.org/10.1073/pnas.2024292118>
- Ruggeri, K., Večkalov, B., Bojanić, L., Andersen, T. L., Ashcroft-Jones, S., Ayacaxli, N., Barea-Arroyo, P., Berge, M. L., Bjørndal, L. D., Bursalioglu, A., Bühler, V., Čadek, M., Çetinçelik, M., Clay, G., Cortijos-Bernabeu, A., Damnjanović, K., Dugue, T. M., Esberg, M., Esteban-Serna, C., . . . Folke, T. (2021). The general fault in our fault lines. *Nature Human Behaviour*, 5(10), 1369–1380. <https://doi.org/10.1038/s41562-021-01092-x>
- Sawaoka, T., & Monin, B. (2018). The paradox of viral outrage. *Psychological Science*, 29(10), 1665–1678.
- Smith, A. (1822). *The theory of moral sentiments* (Vol. 1). J. Richardson.
- Tosi, J., & Warmke, B. (2020). *Grandstanding: The use and abuse of moral talk*. Oxford University Press.
- Waytz, A., Iyer, R., Young, L., Haidt, J., & Graham, J. (2019). Ideological differences in the expanse of the moral circle. *Nature Communications*, 10(1), Article 4389. <https://doi.org/10.1038/s41467-019-12227-0>