



ELSEVIER

Review

Inside the funhouse mirror factory: How social media distorts perceptions of norms

Claire E. Robertson¹, Kareena S. del Rosario¹ and Jay J. Van Bavel²

The current paper explains how modern technology interacts with human psychology to create a funhouse mirror version of social norms. We argue that norms generated on social media often tend to be more extreme than offline norms which can create false perceptions of norms—known as pluralistic ignorance. We integrate research from political science, psychology, and cognitive science to explain *how* online environments become saturated with false norms, *who* is misrepresented online, *what* happens when online norms deviate from offline norms, *where* people are affected online, and *why* expressions are more extreme online. We provide a framework for understanding and correcting for the distortions in our perceptions of social norms that are created by social media platforms. We argue the funhouse mirror nature of social media can be pernicious for individuals and society by increasing pluralistic ignorance and false polarization.

Addresses¹ Psychology Department, New York University, USA² Department of Psychology Center for Neural Science, New York University, Norwegian School of Economics, USACorresponding author: Van Bavel, Jay J (jay.vanbavel@nyu.edu)**Introduction**

The internet is one of the fastest technological revolutions in human history [1]. Almost 5 billion people worldwide use social media, and the average social media user now spends about two and a half hours a day online [2]. Alas, the online environment is far from a true representation of the offline world. In this paper, we argue that social media is akin to a funhouse mirror, reflecting and warping our collective sense of what is normative [3]. When people stare into the mirror they do not see a true version of reality, but instead one that

has been distorted by a small but vocal minority of extreme outliers whose opinions create illusory norms. In turn, these outliers are often amplified by design features and algorithms that prioritize engaging content.

Online discussions are dominated by a surprisingly small, extremely vocal, and non-representative minority. Research on social media has found that, while only 3 % of active accounts are toxic, they produce 33 % of all content [4]. Furthermore, 74 % of all online conflicts are started in just 1 % of communities [5], and 0.1 % of users shared 80 % of fake news [6,7]. Not only does this extreme minority stir discontent, spread misinformation, and spark outrage online, they also bias the meta-perceptions of most users who passively “lurk” online. This can lead to false polarization and pluralistic ignorance, which are linked to a number of problems including drug and alcohol use [8], intergroup hostility [9,10], and support for authoritarian regimes [11]. Furthermore, exposure to extreme content can normalize unhealthy and dangerous behavior. For example, teens exposed to extreme content related to alcohol consumption thought dangerous alcohol consumption was normative [12].

The current paper explains who social media creates a funhouse mirror version of reality. We draw from work in political science, psychology, and cognitive science to explain *how* online environments become saturated with false norms, *who* is misrepresented online, *what* happens when online norms deviate from offline norms, *where* online people are affected, and *why* expressions are more extreme online. We provide a framework for understanding and correcting the distortions in our social perceptions created by social media platforms. We argue the funhouse mirror nature of social media can be pernicious for individuals and society by causing pluralistic ignorance and false polarization.

How do norms become distorted online?

Social norms are defined as the “*predominant behaviors, attitudes, beliefs, and codes of conduct of a group*” [13]. Norms are powerful because they are inherently tied to people’s social identity — the norms one follows signal what group they are in and strengthen their ties with their

Current Opinion in Psychology 2024, 60:101918This review comes from a themed issue on **Norm Change (2024)**Edited by **Cristina Bicchieri**, **Michele Gelfand**, and **Giulia Andrighetto**For complete overview about the section, refer [Generation COVID: Coming of Age Amid the Pandemic \(2024\)](#)

Available online 26 September 2024

<https://doi.org/10.1016/j.copsyc.2024.101918>

2352-250X/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

ingroup [14]. Once a norm has been established within a social group, new group members are especially quick to learn and follow new norms, and those norms are relatively impervious to reinforcement learning [15,16]. Conforming to norms demonstrates commitment to the group, whether that involves supporting group values, policing wrongdoers, or adopting specific worldviews [17–19]. When a norm is ambiguous, such as deciding whether an action is permissible, people base their decisions on the group consensus [20]. Thus, correctly and accurately detecting social norms is critical for social acceptance.

Detecting social norms can be a challenge, as it requires one to attend to the behaviors and opinions of many group members to form a model of how to behave. Thus, rather than encoding and memorizing each individual exemplar of normative behavior and opinion, people instead form an average representation of a series of exemplars in a group via the process of ensemble encoding [21,22]. Ensemble coding is cognitively efficient, allowing people to encode a single representation of a set of stimuli, rather than encoding and memorizing every item [21]. Socially, ensemble coding allows people to form a single estimation of group emotion or opinion, rather than individually encoding each person's reaction [23,24]. Thus, one might gather information about what opinion is normative over repeated interactions with others to form an average representation of a group's opinion. In this way, people encode the social norms from posts and comments in online forums and social media platforms.

Whose opinions are represented online?

While ensemble coding is efficient, it can become distorted online due to the structure of the normative information. False norms emerge, in part, because social media is dominated by a small number of extreme people who post only their most extreme opinions, and do so at a very high volume—often posting dozens of times more than others [25–27], while more moderate or neutral opinions are practically invisible online. Encountering a disproportionate volume of extreme opinions can lead to false perceptions that the norms are far more extreme than they actually are.

This appears to be a general phenomenon online as it occurs across domains and platforms. For instance, online consumer reviews often reflect extremely positive or negative experiences, resulting in ratings that are either perfect or terrible, with little in between [28]. Similarly, on platforms like Instagram, there is a norm to present oneself as interesting, attractive, and successful, showing only the most extremely positive and flattering content to garner peer approval [29]. On LinkedIn, people disproportionately report successes and accomplishments rather than failures [30]. This leads to a feed

where rich models and highly successful people appear to be the norm and can implicitly pressure users to distort, filter, and curate their posts to fit in or gain social status.

In online political discussions, the people who post frequently on social media are often the most ideologically extreme [31,32]. Indeed, 97 % of political posts from Twitter/X come from just 10 % of the most active users on social media, meaning that about 90 % of the population's political opinions are being represented by less than 3 % of tweets online [33]. This is a marked difference from offline polling data showing that most people are ideologically moderate, uninterested in politics, and avoid political discussions when they are able [34–36]. In discussions of the Covid-19 vaccine discussion on Twitter, only 0.35 % of people were in true echo chambers, and yet those users dominated the overall discourse [37]. Similarly, an analysis of social media 448,103 users found that a third of low-credibility posts were shared by just 10 accounts [26] (see also [38]). This renders moderate opinions effectively invisible on social media, leaving the most extreme perspectives most visible for users (see Figure 1).

Receiving such biased inputs from the online environment can lead to extremely biased outputs when people rely on ensemble coding to form representations of the opinions of the general public. This may be especially problematic for topics like politics, where opinions are invisible and people are generally hesitant to share their opinions with others in everyday life. This increases the difficulty of ascertaining true offline norms. People base their perception of norms on an unrepresentative sample of opinions and images leading to a distorted view of social norms. This may be exacerbated by the fact that people tend to weight extreme content more heavily when taking the average of a set of stimuli [23], and assume greater moral outrage from a post than the authors of posts themselves report feeling [39]. Thus, the conceptualized average of opinions for one's ingroup and outgroup in the online world can be far more extreme than true offline norms.

What norms dominate online discourse?

Considering that online discourse is dominated by the most extreme people, it is not surprising that negativity, intergroup hostility, and polarization appear strikingly prevalent online—often exceeding exposure to similar content in the real world and all other forms of media (See Figure 2 [40–43]) The most widely shared content on Twitter/X and Facebook is moralized content, such as expressions of outrage and hostility towards political outgroups [40,42,44]. When positive content does go viral, it also tends to be high valence, inducing emotions like awe [45]. As such, the norms generated on some social media platforms might be more hostile than the offline world.

Figure 1

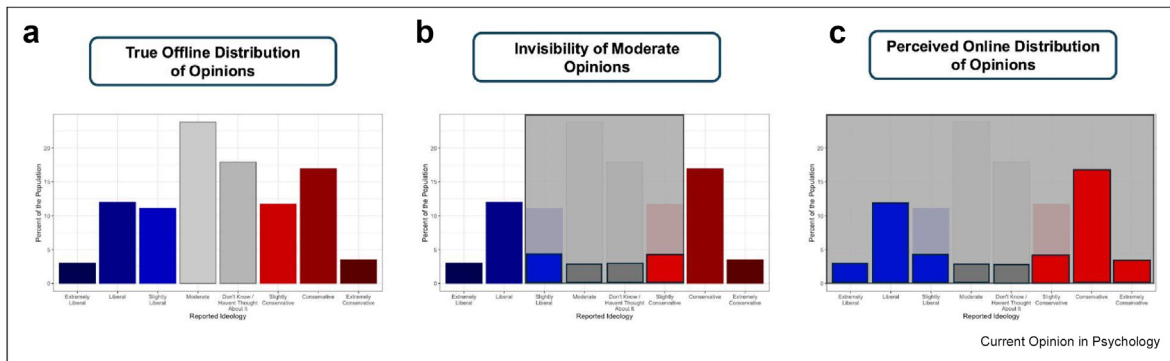
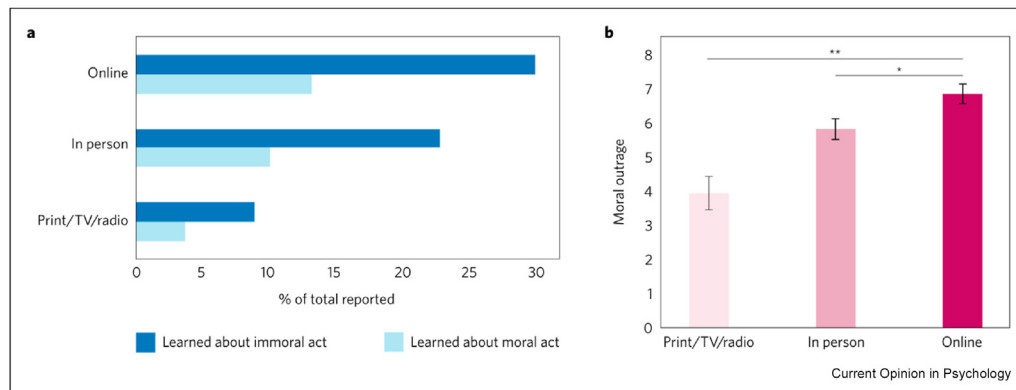


Figure demonstrating how the invisibility of moderate opinion changes the distribution of political opinions online. Data in all panels show data from the American National Election Services from 2016. Participants answered the question “what is your political ideology?” on a likert scale ranging from 1 (Extremely Liberal) to 8 (Extremely Conservative). Responses for “Moderate” or “I don’t know/Haven’t thought about it” are in gray. Panel A shows the distribution of offline opinion as reported from ANES. Panel B shows the gray box representing the “filtering out” of moderate opinions on social media. Panel C shows the perceived distribution of political opinion online after filtering out moderate opinions.

Figure 2



In a sample of North American adults, (a) People were more likely to learn about immoral acts online than in person or via traditional forms of media (print, television, and radio). The figure displays the percentage of total reported moral/immoral acts that were learned about in each setting. (b) Immoral acts encountered online evoked more outrage than immoral acts encountered in person or via traditional forms of media. Error bars represent S.E.M. (Figure adapted from Ref. [41]).

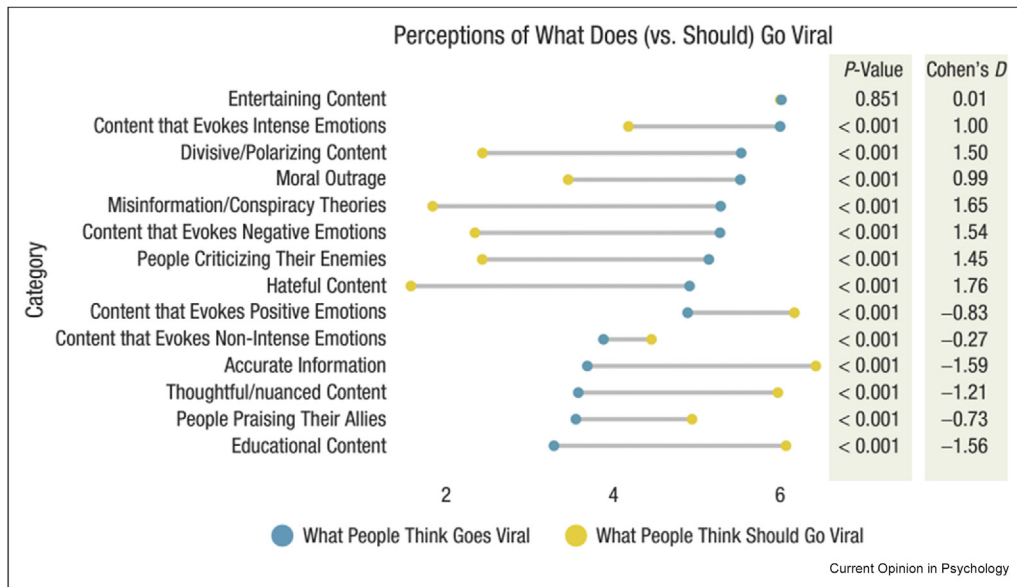
Social media feeds are also curated to normalize extreme norms—from unrealistic beauty standards to outrageous benchmarks for success—thereby fostering a false reality [46]. Even if people recognize that certain visible behaviors (e.g., beauty norms) do not reflect how people *actually* are (i.e., descriptive norm), they are still reinforced through “likes,” signaling what is socially desirable (i.e., prescriptive norm). These extreme norms can leave people feeling inadequate and constant exposure to extreme outliers in body shape on platforms like Instagram may contribute to lower body image and depression in teen girls [47].

Research directly linking distorted online norms with offline norm misperception is still in its infancy, but

early evidence suggests that spillover does occur in some domains (e.g., online content that normalizes alcohol consumption can lead college students to overestimate how common drinking is offline, to endorse drinking behavior, and to engage in drinking behaviors themselves [48–50]). However, it is plausible that some users may recognize the distorted nature of online content and place less weight on this information when calibrating their understanding of norms. We therefore call for more research on the impact of online content on perceptions of real world norms (see also [51]).

Not only are these norms different from offline norms, but they also contradict social media users’ expressed

Figure 3



In a sample of North American adults, there were stark differences between the content that people (n = 511) think goes viral (shown in blue) and the content people think should go viral (shown in yellow). Questions were answered on a scale from 1 (strongly disagree) to 7 (strongly agree); 4 is the midpoint. The p-value column represents p values from paired (within subjects) *t* tests and Cohen's D represents effect sizes of the discrepancy. (Figure adapted from Ref. [52]).

preferences. While most users acknowledge that negative and extreme content is most prevalent online, the majority of users say they would prefer more positive and nuanced content (see Figure 3 [52]). This discrepancy suggests that these false norms make divisiveness seem more pervasive than it actually is—and certainly more than people want it to be. Even though incivility from politicians is increasing and is increasingly socially rewarded online [53], people report that they want to hear less from uncivil politicians [54]. So, if most people say they want positive content, what causes the norms of negativity and extremity to be so prevalent online?

Why are false norms worse online than offline?

Social media operates in an attention economy, where design features and algorithms are designed to elicit as much engagement as possible [3,55]. Platforms then sell ad space to companies based on indices of attention. As such, there is a strong incentive for users to create content that captures attention and maximizes engagement—rather than content that reflects reality. Given that users who are the most active on social media are also the most extreme [25,32]), this creates a perverse incentive structure to reward surprising, negative, extreme or divisive content. For instance, news stories that express outgroup animosity are 67 % more likely to be shared on social media [42]. Thus, people with more extreme or hostile beliefs tend to dominate discourse — drowning out or overshadowing more mild or nuanced

content [31] and leading to false beliefs about the norms of a community.

These online dynamics are amplified by the design features and recommendation algorithms on various platforms [56,57]. For instance, a recent analysis of the algorithm on Twitter/X found that it prioritizes evocative content [44]. This incentivizes users to create this type of content which can help them build a large following while warping public perceptions of norms. Indeed, people who are focused on gaining social status are the most hostile online [25]. These strivers may further distort norms as they rise in status.

This is compounded due to the fact that there is often little motivation for someone to post a nuanced or moderate opinion on social media. Moreover, nuanced or moderate posts often risk hostility from more extreme ingroup and outgroup members, especially since such hostility has little cost for the aggressor due to the social distance the online environment affords [58]. Indeed, people who are politically moderate were more likely to report being harassed online, even though they were also less likely to post [31]. People who hold less extreme beliefs have less investment into arguing, and when attacked by people who have more strongly held beliefs, perceive it as more hostile. The fact that people who “troll” other people typically have higher dark triad characteristics also does not encourage nuanced debate [59].

Conclusion

Every day users casually scroll through an estimated 300 feet of newsfeed on social media—roughly the height of the Statue of Liberty [60]. As they casually scroll through this content, they are forming beliefs about the state of the world as well as inferences about the beliefs of members of their own social network and community. But these inferences are often based on the most extreme voices. Being overexposed to the most extreme opinions from the most extreme people can have real consequences. Believing that one's political outgroup endorses extreme political positions may lead to biased meta-perceptions, pluralistic ignorance, and false polarization [9,10,39,61].

This poses a pernicious problem for society: how do people differentiate what is normative vs. unpopular when the content that drives the most engagement is often from a minority of extreme (and often hostile) users? This is an especially challenging problem since these misperceptions might be driven by factual content—making them uniquely difficult to address through content moderation and other mechanisms designed to root out misinformation. Yet, if we are unable to solve this problem, people may develop a distorted sense of reality as we rely on the funhouse mirror to reflect the truth.

Credit author statement

Claire E. Robertson: Conceptualization, Writing – Original Draft, Writing – Review and Editing, Visualization.

Kareena S. Del Rosario: Conceptualization, Writing – Original Draft, Writing – Review and Editing.

Jay J. Van Bavel: Conceptualization, Writing – Original Draft, Writing – Review and Editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jay Van Bavel reports financial support was provided by Google Jigsaw. Jay Van Bavel reports financial support was provided by John Templeton World Charity Foundation (TWCF-2022-30561). If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

References of particular interest have been highlighted as:

** of outstanding interest

- Xu M, David JM, Kim SH: **The fourth industrial revolution: opportunities and challenges.** *Int J Financ Res* Feb. 2018, **9**:90, <https://doi.org/10.5430/ijfr.v9n2p90>.
- Petrosyan A: **“Internet and social media users in the world 2022,”** *Statista* [Online]. Available: <https://www.statista.com/statistics/617136/digital-population-worldwide/>. Accessed 30 September 2022.
- Fisher M: *The chaos machine: the inside story of how social media rewired our minds and our world.* New York: Little, Brown and Company; 2022.
- Kumar D, Hancock J, Thomas K, Durumeric Z: **Understanding the behaviors of toxic accounts on reddit.** In *Proceedings of the ACM web conference 2023, in WWW '23.* New York, NY, USA: Association for Computing Machinery; Apr. 2023:2797–2807, <https://doi.org/10.1145/3543507.3583522>.
- Kumar S, Hamilton WL, Leskovec J, Jurafsky D: **Community interaction and conflict on the web.** In *Proceedings of the 2018 world wide web conference, in WWW '18. Republic and canton of Geneva, CHE: international world wide web conferences steering committee;* Apr. 2018:933–943, <https://doi.org/10.1145/3178876.3186141>.
- Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D: **Fake news on Twitter during the 2016 U.S. presidential election.** *Science* Jan. 2019, **363**:374–378, <https://doi.org/10.1126/science.aau2706>.
- Guess A, Nagler J, Tucker J: **Less than you think: prevalence and predictors of fake news dissemination on Facebook.** *Sci Adv* Jan. 2019, **5**:eaau4586, <https://doi.org/10.1126/sciadv.aau4586>.
- Prentice DA, Miller DT: **Pluralistic ignorance and alcohol use on campus: some consequences of misperceiving the social norm.** *J Pers Soc Psychol* 1993, **64**:243–256, <https://doi.org/10.1037/0022-3514.64.2.243>.
- Lees J, Cikara M: **Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts.** *Nat Human Behav* Mar. 2020, **4**, <https://doi.org/10.1038/s41562-019-0766-4>. Art. no. 3.
- Moore-Berg SL, Ankori-Karlinsky L-O, Hameiri B, Bruneau E: **Exaggerated meta-perceptions predict intergroup hostility between American political partisans.** *Proc Natl Acad Sci USA* Jun. 2020, **117**:14864–14872, <https://doi.org/10.1073/pnas.2001263117>.
- Kuran T: *Private truths, public lies: the social consequences of preference falsification.* Harvard University Press; 1998.
- Davis J, Desmarais S, Giguère B: **Under the influence: how viewing extreme partying and drinking on social media shapes group perceptions.** *J Soc Psychol* 2023:1–14, <https://doi.org/10.1080/00224545.2023.2219384>. vol. 0, no. 0.
- Cialdini RB, Jacobson RP: **Influences of social norms on climate change-related behaviors.** *Curr. Opin. Behav. Sci.* Dec. 2021, **42**:1–8, <https://doi.org/10.1016/j.cobeha.2021.01.005>.
- Van Bavel JJ, Packer DJ: *The power of us: harnessing our shared identities to improve performance, increase cooperation, and promote social harmony.* New York Boston London: Little, Brown Spark; 2021.
- Brady WJ, McLoughlin K, Doan TN, Crockett MJ: **How social learning amplifies moral outrage expression in online social networks.** *Sci Adv* Aug. 2021, **7**:eabe5641, <https://doi.org/10.1126/sciadv.abe5641>.

16. Ho MK, MacGlashan J, Littman ML, Cushman F: **Social is special: a normative framework for teaching with and learning from evaluative feedback.** *Cognition* 2017, **167**:91–106, <https://doi.org/10.1016/j.cognition.2017.03.006>.
17. Atari M, et al.: **Morally homogeneous networks and radicalism.** *Soc Psychol Personal Sci* Aug. 2022, **13**:999–1009, <https://doi.org/10.1177/19485506211059329>.
18. Goldenberg A, et al.: **Homophily and acrophily as drivers of political segregation.** *Nat Human Behav* Nov. 2023:1–12, <https://doi.org/10.1038/s41562-022-01474-9>.
19. Kennedy B, et al.: **The (moral) language of hate.** *PNAS Nexus* Jul. 2023, **2**:pgad210, <https://doi.org/10.1093/pnasnexus/pgad210>.
20. Cornwell JFM, Jago CP, Higgins ET: **When group influence is more or less likely: the case of moral judgments.** *Basic Appl Soc Psychol* Nov. 2019, **41**:386–395, <https://doi.org/10.1080/01973533.2019.1666394>.
21. Alvarez GA: **Representing multiple objects as an ensemble enhances visual cognition.** *Trends Cognit Sci* Mar. 2011, **15**:122–131, <https://doi.org/10.1016/j.tics.2011.01.003>.
22. Parkes L, Lund J, Angelucci A, Solomon JA, Morgan M: **Compulsory averaging of crowded orientation signals in human vision.** *Nat Neurosci* Jul. 2001, **4**, <https://doi.org/10.1038/89532>. Art. no. 7.
23. Goldenberg A, et al.: **Amplification in the evaluation of multiple emotional expressions over time.** *Nat Human Behav* Jun. 2022, **6**:1408–1416, <https://doi.org/10.1038/s41562-022-01390-y>.
24. Haberman J, Whitney D: **Rapid extraction of mean emotion and gender from sets of faces.** *Curr Biol* Sep. 2007, **17**:R751–R753, <https://doi.org/10.1016/j.cub.2007.06.039>.
25. Bor A, Petersen MB: **The psychology of online political hostility: a comprehensive, cross-national test of the mismatch hypothesis.** *Am Polit Sci Rev* Feb. 2022, **116**:1–18, <https://doi.org/10.1017/S0003055421000885>.
26. DeVerna MR, Aiyappa R, Pacheco D, Bryden J, Menczer F: **Identifying and characterizing superspreaders of low-credibility content on Twitter.** *PLoS One* May 2024, **19**, e0302201, <https://doi.org/10.1371/journal.pone.0302201>.
27. Heltzel G, Laurin K: **Why Twitter sometimes rewards what most people disapprove of: the case of cross-party political relations.** *OSF*; May 03, 2024, <https://doi.org/10.31234/osf.io/f3p4s>.
28. Schoenmueller V, Netzer O, Stahl F: **The polarity of online reviews: prevalence, drivers and implications.** *J Mar Res* Oct. 2020, **57**:853–877, <https://doi.org/10.1177/0022243720941832>.
29. Yau JC, Reich SM: **It's just a lot of work': adolescents' self-presentation norms and practices on Facebook and Instagram.** *J Res Adolesc* 2019, **29**:196–209, <https://doi.org/10.1111/jora.12376>.
30. van Dijck J: **You have one identity': performing the self on Facebook and LinkedIn.** *Media Cult Soc* Mar. 2013, **35**:199–215, <https://doi.org/10.1177/0163443712468605>.
31. Bail C: *Breaking the social media prism: how to make our platforms less polarizing. First paperback edition.* Princeton: Princeton University Press; 2022.
32. Barberá P, Rivero G: **Understanding the political representativeness of twitter users.** *Soc Sci Comput Rev* Dec. 2015, **33**:712–729, <https://doi.org/10.1177/0894439314558836>.
33. Hughes A: **"A small group of prolific users account for a majority of political tweets sent by U.S. adults,"** *Pew Research Center* [Online]. Available: <https://www.pewresearch.org/short-reads/2019/10/23/a-small-group-of-prolific-users-account-for-a-majority-of-political-tweets-sent-by-u-s-adults/>. Accessed 16 November 2023.
34. American National Election Studies: *ANES 2020 time series study full release* [dataset and documentation]. Jul. 19, 2021 [Online]. Available: www.electionstudies.org.
35. Converse PE: **The nature of belief systems in mass publics (1964).** *Crit Rev* Jan. 1964, **18**:1–74, <https://doi.org/10.1080/08913810608443650>.
36. Krupnikov Y: *The other divide, new edition.* Cambridge: Cambridge University Press; 2022.
37. Pratelli M, Saracco F, Petrocchi M: **Entropy-based detection of Twitter echo chambers.** *PNAS Nexus* May 2024, **3**:pgae177, <https://doi.org/10.1093/pnasnexus/pgae177>.
38. Baribi-Bartov S, Swire-Thompson B, Grinberg N: **Supersharers of fake news on Twitter.** *Science* May 2024, **384**:979–982, <https://doi.org/10.1126/science.adl4435>.
39. Brady WJ, McLoughlin KL, Torres MP, Luo KF, Gendron M, Crockett MJ: **Overperception of moral outrage in online social networks inflates beliefs about intergroup hostility.** *Nat Human Behav* Jun. 2023, **7**:917–927, <https://doi.org/10.1038/s41562-023-01582-0>.
40. Brady WJ, Wills JA, Jost JT, Tucker JA, Bavel JJV: **Emotion shapes the diffusion of moralized content in social networks.** *Proc Natl Acad Sci USA* Jul. 2017, **114**:7313–7318, <https://doi.org/10.1073/pnas.1618923114>.
41. Crockett MJ: **Moral outrage in the digital age.** *Nat Human Behav* Nov. 2017, **1**:769–771, <https://doi.org/10.1038/s41562-017-0213-3>.
42. Rathje S, Van Bavel JJ, van der Linden S: **Out-group animosity drives engagement on social media.** *Proc Natl Acad Sci USA* Jun. 2021, **118**, e2024292118, <https://doi.org/10.1073/pnas.2024292118>.
43. Robertson CE, Pröllochs N, Schwarzenegger K, Pärnamets P, Van Bavel JJ, Feuerriegel S: **Negativity drives online news consumption.** *Nat Human Behav* Mar. 2023:1–11, <https://doi.org/10.1038/s41562-023-01538-4>.
44. Milli S, Carroll M, Pandey S, Wang Y, Dragan AD: *Twitter's algorithm: amplifying anger, animosity, and affective polarization.* arXiv: arXiv:2305.16941. May 26, 2023 [Online]. Available: <http://arxiv.org/abs/2305.16941>. Accessed 7 July 2023.
45. Berger J, Milkman KL: **What makes online content viral?** *J Mar Res* Apr. 2012, **49**:192–205, <https://doi.org/10.1509/jmr.10.0353>.
46. Mun IB, Kim H: **Influence of false self-presentation on mental health and deleting behavior on Instagram: the mediating role of perceived popularity.** *Front Psychol* 2021, **12**, <https://doi.org/10.3389/fpsyg.2021.660484>.
47. Kelly Y, Zilanawala A, Booker C, Sacker A: **Social media use and adolescent mental health: findings from the UK millennium cohort study.** *eClinicalMedicine* Dec. 2018, **6**:59–68, <https://doi.org/10.1016/j.eclinm.2018.12.005>.
48. Vannucci A, Simpson EG, Gagnon S, Ohannessian CM: **Social media use and risky behaviors in adolescents: a meta-analysis.** *J Adolesc* Feb. 2020, **79**:258–274, <https://doi.org/10.1016/j.adolescence.2020.01.014>.
49. Boyle SC, Smith DJ, Earle AM, LaBrie JW: **What 'likes' have got to do with it: exposure to peers' alcohol-related posts and perceptions of injunctive drinking norms.** *J Am Coll Health* May 2018, **66**:252–258, <https://doi.org/10.1080/07448481.2018.1431895>.
50. Boyle SC, LaBrie JW, Froidevaux NM, Witkovic YD: **Different digital paths to the keg? How exposure to peers' alcohol-related social media content influences drinking among male and female first-year college students.** *Addict Behav* Jun. 2016, **57**:21–29, <https://doi.org/10.1016/j.addbeh.2016.01.011>.
51. Brady WJ, Crockett MJ: **Norm psychology in the digital age: how social media shapes the cultural evolution of normativity.** *Perspect Psychol Sci* Jan. 2024, **19**:62–64, <https://doi.org/10.1177/17456916231187395>.
52. Rathje S, Robertson CE, Brady WJ, Van Bavel JJ: **People think that social media platforms do (but should not) amplify divisive content.** *Perspect Psychol Sci* 2024, **19**:781–795.
53. Frimer JA, et al.: **Incivility is rising among American politicians on twitter.** *Soc Psychol Personal Sci* Mar. 2023, **14**:259–269, <https://doi.org/10.1177/19485506221083811>.

54. Feinberg M, Frimer JA: **Incivility diminishes interest in what politicians have to say.** *Soc Psychol Personal Sci* Sep. 2023, **14**:787–795, <https://doi.org/10.1177/19485506221136182>.
55. Van Bavel JJ, Robertson CE, del Rosario K, Rasmussen J, Rathje S: **Social media and morality.** *Annu Rev Psychol* 2024, **75**: 311–340, <https://doi.org/10.1146/annurev-psych-022123-110258>.
56. Brady WJ, Crockett MJ, Van Bavel JJ: **The MAD model of moral contagion: the role of motivation, attention, and design in the spread of moralized content online.** *Perspect Psychol Sci* Jul. 2020, **15**:978–1010, <https://doi.org/10.1177/1745691620917336>.
57. Brady WJ, Jackson JC, Lindström B, Crockett MJ: **Algorithm-mediated social learning in online social networks.** *Trends Cognit Sci* Oct. 2023, **27**:947–960, <https://doi.org/10.1016/j.tics.2023.06.008>.
58. Robertson C, Shariff A, Van Bavel JJ: *Morality in the anthropocene: the perversion of compassion and punishment in the online world.* Feb. 04, 2024, <https://doi.org/10.31234/osf.io/ns34y>.
59. Brubaker PJ, Montez D, Church SH: **The power of schadenfreude: predicting behaviors and perceptions of trolling among reddit users.** *Soc. Media Soc.* Apr. 2021, **7**, 20563051211021382, <https://doi.org/10.1177/20563051211021382>.
60. L. Morant, "The Truth Behind 6 Second Ads," Medium. Accessed: November. 8, 2023. [Online]. Available: <https://medium.com/@Lyndon/the-tyranny-of-six-seconds-592b94160877>.
61. Ploger G: **Polarization all the way down: how coverage of elite and partisan polarization spills over to perceptions of the**

U.S. Mass public. *Polit Commun* 2024, **0**:1–20, <https://doi.org/10.1080/10584609.2024.2312867>.

Further Information on References of Particular Interest

23. Across 9 studies (N = 1583), the authors investigated how ^{**} ensemble coding can be biased by extreme exemplars during iterative exposure to emotional faces. The authors found that, when extreme emotional faces were included, perceivers tended to overestimate the overall emotionality of the sequence of faces.
25. Across 8 studies (total N = 8434), the authors investigated the ^{**} mismatch hypothesis to discover whether the online environment made people more hostile. They found that people who were hostile in online political discourse were equally hostile offline, but hypothesized that online political discussions feel more hostile because aggressors are more visible online.
39. The authors collected data from social media who posted morally ^{**} outrageous content and asked them how morally outraged they actually felt. Across three studies, the authors found that observers systematically overestimated the amount of outrage the authors actually felt at the time they posted their message.
52. In a nationally representative sample of U.S. adults, the authors ^{**} found that, while people report perceiving moral outrage, negativity, misinformation and divisive content as likely to go viral, people reported believing that more positive content such as accurate content, nuanced content, and educational content should go viral online.