Social media, extremism, and radicalization

Aaron Shaw

Fears that YouTube recommendations radicalize users are overblown, but social media still host and profit from dubious and extremist content.

In the U.S., the sentencing of the 2018 Pittsburgh Tree of Life synagogue shooter and the arraignment of the former president for his role in the conspiracy and riots of 6 January 2021 dominated the news in early August 2023. As both incidents were partly inspired, planned, and documented in extremist networks on social media, public reflections about such events should rekindle questions about the role of social media in extremist radicalization in American public life.

Many would blame social media platforms—in particular, their algorithms that sort and recommend content—for the spread of extremist ideas. However, empirical evidence, including a study of YouTube led by Annie Y. Chen in this issue of *Science Advances*, reveals a more complex reality (*1*). The platforms and their algorithms rarely recommend extremist content, yet they remain powerful tools for those who hold extremist beliefs. Radicalized users can still use social media to access and disseminate ideas, build solidarity, or plan and publicize egregious acts. Indeed, despite efforts to remove or reduce the visibility of extremist content, social media platforms like YouTube continue to provide a hospitable environment for content espousing violence, hate, and conspiracist thinking of various kinds (Fig. 1).

Critical accounts of the ills wrought by social media have become commonplace, but the details are still important. YouTube, launched in 2005 and acquired by Google (now Alphabet) in 2006, is one of the most popular social media platforms in the United States (*2*). YouTube's recommendation algorithms, which drive massive amounts of content consumption on the site, have a notorious reputation for surfacing hate speech, unfounded rumors, misinformation, hoaxes, and conspiracies. The platform's recommendations, so the story goes, expose casual users to extremist content, nudging them down "rabbit holes" of (usually right wing) radicalization.

The rabbit holes narrative gained traction in the wake of the 2016 U.S. presidential election. Breathless observers, exemplified by a 2018 op-ed by Zeynep Tufekci in *The New York Times* (*3*), proclaimed YouTube "the great radicalizer" and argued that its recommendation tools "may be one of the most powerful radicalizing instruments of the 21st century." This *Alice in Wonderland* vision of innocents undergoing extremist-contagion-via-algorithms reflects a mix of fantasy and fear typical of moral panics over technology. Such talk also entails a crude, outdated theory of direct media effects. Just exposing someone to media that espouse far-out ideas is unlikely to change their perspective. Deeply held views are not like air-borne illnesses that spread in a few breaths. Rather, contagions of behavior and beliefs are complex, requiring reinforcement to catch on (*4*).

Nevertheless, the visions of rabbit holes, supported by little more than anecdotes, may have been more accurate prior to 2019, when YouTube introduced changes to render extremist content less visible (*5*). Scant empirical evidence was published until several years later (YouTube held but did not release data that could have supported independent tests). Other platforms, including Facebook, Reddit, and X (formerly Twitter), likewise expanded moderation strategies in response to mounting criticism for their role in hosting hate speech, incivility, and worse. Many of the interventions reduced hateful and dangerous content [e.g., (*6*)]. In other words, the social media platforms and society are far from helpless in the face of an upsurge of hateful and uncivil content

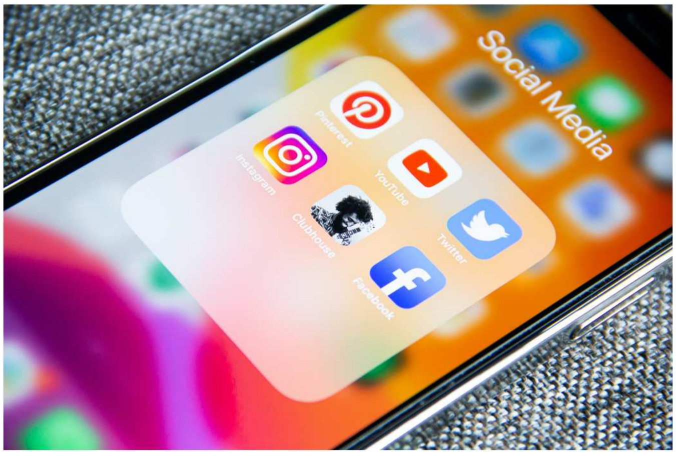

Fig. 1. Turbulent times for social media. A new study by Annie Chen *et al.* disentangles the relationships between online behavior and prior beliefs. The study confirms that platforms like YouTube can, and should, do much more to restrict the reach of extremist content to the dedicated audiences that seek it out. Photo by Adem AY on Unsplash

Department of Communication Studies, Northwestern University, Evanston, IL, USA. Email: aaronshaw@northwestern.edu

but, instead, are increasingly well equipped to identify it and minimize its reach.

The new study led by Chen in collaboration with Brendan Nyhan, Jason Reifler, Ronald E. Robertson, and Christo Wilson finds that exposure to alternative and extremist YouTube videos happens among users who already hold resentful attitudes about race and gender and who seek out this content via channel subscriptions and referrals from other sites. Such sites include fringe social media platforms like Parler and Gab, both of which embrace radically permissive content policies and extremist political movements. By contrast, algorithmic recommendations within YouTube generate a very small amount of the traffic to alternative and extremist content. In this last respect, the findings echo a recent piece in *Proceedings of the National Academy of Sciences* that first documented the scarcity of rabbit hole events among YouTube users (*7*). A key contribution of Chen and colleagues consists of matched survey and web browsing data, which allows them to disentangle the relationships between online behavior and prior beliefs. Doing so shows that consumers of alternative and extremist content previously espoused extremist beliefs. The study cannot rule out the possibility that these individuals acquired their extremist views via YouTube recommendations prior to 2019, but, at some point, we should recall that violent extremism has a deeply entrenched history in American society that pre-dates social media [e.g., (*8*)].

The fact that a substantial proportion of the consumers of extremist content on YouTube arrived from other, extremist sites also speaks to a distinct, pernicious—and empirically documented—role of social media in the contemporary epidemic of extremist violence in the U.S. Participation in extremist online spaces correlates with increased participation in subsequent incidents of extremist civil unrest (*9*). Thankfully, most such incidents are neither mass shootings nor electoral malfeasance, but both help illustrate the pattern. The Tree of Life shooter appears to have engaged with violent antisemitic groups online. The perpetrators of the January 6 debacle coordinated across various platforms and communities. Engaging with and contributing to communities of like-minded extremists may not have caused these individuals to adopt such radicalized beliefs in the first place, but the social support that they found online may have catalyzed them to adopt even more extreme views and to take actions they once might have considered taboo. Future research should continue to pull at these threads.

Meanwhile, the terrain of social media use and governance remains fraught. Online ecosystems have fragmented as younger users and others have congregated in newer platforms like TikTok or decentralized environments like Mastodon. More polarized and more misinformation-suffused right wing media sources had greater visibility and engagement on Facebook around the 2020 election (*10*). Platform safeguards put in place around the 2020 election to prevent the spread of misinformation online have been weakened ahead of 2024. Elon Musk has dismantled most of the trust and safety infrastructure of X and appears to hold deeper commitments to extremist speech than civility. Republican Ohio Representative and House Judiciary Committee Chair Jim Jordan has launched a burdensome, evidence-optional inquisition into the conduct of social media companies and academic researchers who sought to protect electoral integrity in 2020. Jordan has targeted, among others, Kate Starbird of the University of Washington, whose primary faults seem to have been working for over a decade to identify dangerous rumors in social media and sharing findings with interested parties (*11*). The adoption of large language models and generative AI tools will bring new challenges and disruptions.

The turbulent context is part of what makes Chen and colleagues' work important. The science of algorithmic recommendation systems, content moderation, and digital media must continue to evolve quickly. We must continue to investigate the means by which ideas that threaten public safety and institutional integrity spread, take hold, and endanger lives.

The platforms present a moving target. Just because they do not incidentally expose visitors to radical extremist content today does not mean that they never did or that they will not do so again. Furthermore, Chen and colleagues' study confirms that platforms like YouTube can, and should, do much more to restrict the reach of extremist content to the dedicated audiences that seek it out. At a minimum, YouTube and its parent Alphabet should divest from revenue generating activities related to content that contradicts their public commitments (*12*) to reduce the spread of hate speech, harassment, and harmful conspiracy theories.

References

1. A. Chen, B. Nyhan, J. Reifler, R. E. Robertson, C. Wilson, Subscriptions and external links help drive resentful users to alternative and extremist YouTube videos. *Sci. Adv.* **9**, eadd8080 (2023).
2. B. Auxier, M. Anderson, "Social media use in 2021" (Pew Research Center, 2021); www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/.
3. Z. Tufekci, "Opinion | YouTube, the great radicalizer," *The New York Times*, 10 March 2018.
4. D. Centola, *How Behavior Spreads: The Science of Complex Contagions*. (Princeton University Press, 2018).
5. The YouTube Team, "Continuing our work to improve recommendations on YouTube" (blog.youtube, 2019); https://blog.youtube/news-and-events/continuing-our-work-to-improve/.
6. M. H. Ribeiro, S. Jhaver, S. Zannettou, J. Blackburn, G. Stringhini, E. De Cristofaro, Do platform migrations compromise content moderation? Evidence from r/The_Donald and r/Incels. *Proc. ACM Hum.-Comput. Interact.* **5**, 1–24 (2021).
7. H. Hosseinmardi, A. Ghasemian, A. Clauset, M. Mobius, D. M. Rothschild, D. J. Watts, Examining the consumption of radical content on YouTube. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2101967118 (2021).
8. K. Belew, *Bring the War Home: The White Power Movement and Paramilitary America* (Harvard Univ. Press, 2019).
9. D. Karell, A. Linke, E. Holland, E. Hendrickson, "Born for a storm": Hard-right social media and civil unrest. *Am. Sociol. Rev.* **88**, 322–349 (2023).
10. S. González-Bailón, D. Lazer, P. Barberá, M. Zhang, H. Allcott, T. Brown, A. Crespo-Tenorio, D. Freelon, M. Gentzkow, A. M. Guess, S. Iyengar, Y. M. Kim, N. Malhotra, D. Moehler, B. Nyhan, J. Pan, C. V. Rivera, J. Settle, E. Thorson, R. Tromble, A. Wilkins, M. Wojcieszak, C. K. de Jonge, A. Franco, W. Mason, N. J. Stroud, J. A. Tucker, Asymmetric ideological segregation in exposure to political news on Facebook. *Science* **381**, 392–398 (2023).
11. S. L. Myers, S. Frenkel, "G.O.P. targets researchers who study disinformation ahead of 2024 election," *The New York Times*, 19 June 2023.
12. YouTube, "Our commitments: Hate speech & harassment policy - How YouTube works" (2023); www.youtube.com/howyoutubeworks/our-commitments/standing-up-to-hate/.

10.1126/sciadv.adk2031