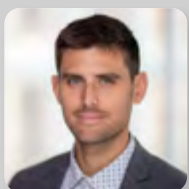




WHAT WE TALK ABOUT WHEN WE TALK ABOUT DEPLATFORMING TRUMP



BY
SCOTT BABWAH BRENNEN



&
MATT PERAULT

Scott Babwah Brennen is head of online expression policy at the Center on Technology Policy at UNC-Chapel Hill. Matt Perault is the director of the Center on Technology Policy at UNC-Chapel Hill, a professor of the practice at UNC's School of Information and Library Science, and a consultant on technology policy issues at Open Water Strategies.

THE TAMING OF INTERNET PLATFORMS - A LOOK AT THE EUROPEAN DIGITAL SERVICES ACT

By Natascha Just



WHY TECH REGULATORS NEED TO THINK LIKE GOOGLE

By Terry Flew



WHAT WE TALK ABOUT WHEN WE TALK ABOUT DEPLATFORMING TRUMP

By Scott Babwah Brennen & Matt Perault



AESTHETICS, TECHNOLOGY, AND REGULATIONS

By Gregory Day



RANDOMIZED EXPERIMENTS FOR ONLINE CONTENT MODERATION POLICY

By Imanol Ramírez



PLATFORM CONTENT STANDARDIZATION

By Marta Cantero Gamito



WHAT WE TALK ABOUT WHEN WE TALK ABOUT DEPLATFORMING TRUMP

By Scott Babwah Brennen & Matt Perault

Following his role in the January 6th attacks on the Capitol, several online platforms banned President Donald Trump from their platforms. While scholars and commentators have spent 18 months debating the value of removing Trump from major online platforms, we still lack sufficient empirical analysis of the positive and negative consequences the ban has had on public discourse, extremism, and Trump's power and influence. But before we can assess the costs and benefits of removal, we first must develop a common understanding of the metrics we might use to evaluate the impacts of the ban. When we talk about whether deplatforming worked or failed, what do we mean by "worked" and what do we mean by "failed"? This article provides 13 metrics that analysts or researchers could consider as a means of evaluating the impacts of the ban. While analyzing any of these will involve overcoming serious challenges, from data access, to resources, to how to attribute observed changes narrowly to Trump's removal, identifying what metrics we should consider is an important first step to understanding the impact of banning Trump.

Visit www.competitionpolicyinternational.com for access to these articles and more!

Scan to Stay Connected!

Scan here to subscribe to CPI's **FREE** daily newsletter.



01

INTRODUCTION

Following his role in the January 6th attacks on the Capitol, a series of online platforms, including Twitter, YouTube, Facebook, and Instagram, banned President Donald Trump from their platforms. Eighteen months later, as the midterms and the next presidential campaign approach, Twitter might have new ownership, and Facebook's ban on Trump is set to expire in January 2023, there is once again a national discussion about the value of removing Trump from major online platforms.

In a post explaining Facebook's decision to remove Trump in early 2021, Mark Zuckerberg wrote, “[w]e removed these statements yesterday because we judged that their effect – and likely their intent – would be to provoke further violence.”² Rather than arguing that Trump expressly violated community standards, Zuckerberg justified the ban by appealing to the likely effect of allowing Trump to remain on the platform.

But what were the “effects” of removing Trump from the platforms? Did it decrease the amount or reach of problematic content? Did it encourage followers to seek out more radical communities on alternative platforms? Did it buttress or undermine Trump's political power?

Empirical understanding of the effects of policy decisions can and should be used to design and target policy interventions more efficiently and effectively. Any consideration of whether to reinstitute Trump to major platforms should be based, at least in part, on a rigorous examination of the good and bad consequences the ban has had on public discourse, on extremism, and on Trump's power and influence. The value of this exercise extends beyond a decision on Trump's account. When platforms face decisions in the future on whether or not to remove users or sitting government officials, those decisions should be informed by an assessment of similar interventions in the past.

But before we can assess the costs and benefits of removal, we first must develop a common understanding of

the metrics we might use to evaluate the impacts of the ban. When we talk about whether deplatforming worked or failed, what do we mean by “worked” and what do we mean by “failed”?

Toward this end, we provide 13 metrics that analysts or researchers could consider as a means of evaluating the impacts of the ban. While analyzing any of these will involve overcoming serious challenges, from data access, to resources, to how to attribute observed changes narrowly to Trump's removal, identifying what metrics we should consider is an important first step to understanding the impact of banning Trump.

02

WHAT WE KNOW ABOUT DEPLATFORMING

Recently, scholars have offered some empirical insight into the effects of removing individuals or communities from major platforms. Broadly speaking, analysis suggests that deplatforming reduces the amount of prohibited and/or problematic content on the platform from which the user or community is banned. Analyses of the removal of prominent influencers on Twitter (Jhaver et al., 2021)³ and of communities on Reddit (Chandrasekharan et al., 2017)⁴ both observed declines in related conversations, the activity of followers, and the toxicity of content. In contrast, observing conspiracy communities on Facebook, Innes & Innes (2021) found that “minion accounts” worked to “replatform” and share the content that had once been shared by now-banned accounts and groups, meaning that much of the problematic content continued to circulate.⁵

There is also evidence that, in many cases, followers of banned individuals or communities moved to alternative platforms. Notably, when users move to alternative platforms, many become more active, posting more content (Ali et al., 2021)⁶ — at least initially (Rauchfleisch & Kaiser,

2 <https://about.fb.com/news/2021/01/responding-to-the-violence-in-washington-dc/>.

3 Jhaver, S., Boylston, C., Yang, D., & Bruckman, A. (2021). Evaluating the effectiveness of deplatforming as a moderation strategy on Twitter. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1–30.

4 Chandrasekharan, E., Pavalanathan, U., Srinivasan, A., Glynn, A., Eisenstein, J., & Gilbert, E. (2017). You Can't Stay Here: The Efficacy of Reddit's 2015 Ban Examined Through Hate Speech. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW), 31:1-31:22. <https://doi.org/10.1145/3134666>.

5 Innes, H., & Innes, M. (2021). De-platforming disinformation: Conspiracy theories and their control. *Information, Communication & Society*, 1–19.

6 Ali, S., Saeed, M. H., Aldreabi, E., Blackburn, J., De Cristofaro, E., Zannettou, S., & Stringhini, G. (2021). Understanding the effect of deplatforming on social networks. *13th ACM Web Science Conference 2021*, 187–195.

2021).⁷ In some cases, users' content on alternative platforms is more toxic than on primary platforms (Ali et al, 2021; Horta Ribeiro, 2021).⁸ However, the reach of this more toxic and more abundant content is significantly lower (Rauchfleisch & Kaiser, 2021; Ali et al., 2021; Horta Ribeiro, 2021) — the audience on alternative platforms simply cannot replace that which was lost on mainstream sites. Importantly, however, existing analyses only examine the movement of users to single alternative platforms, they do not capture activity spread across multiple platforms.

Existing analyses provide less specific insight into the impacts of deplatforming Trump from major platforms. Most studies focus narrowly on how the ban impacted problematic content. An analysis by the for-profit Signal Labs covered by the Washington Post (Dwoskin & Timberg, 2021)⁹ found that misinformation about the 2020 presidential election on Twitter reduced by 73 percent after Trump was banned from the platform. However, a *New York Times* analysis (Alba et al., 2021)¹⁰ found that after Trump was banned from major platforms, a handful of his statements still eventually received as many likes or shares as his posts before the ban. Similar to Innes & Innes (2021), the article observes that “Mr. Trump’s most ardent supporters continue to spread his message — doing the work that he had been unable to do himself.”

Given the narrowness of both the literature on the impact of deplatforming in general and of deplatforming Trump specifically, we need more, broader analyses of the range of potential impacts of the ban across platforms and across media. But in order to produce those analyses, we first need a shared understanding of what we mean when we talk about “impact.”

“There is also evidence that, in many cases, followers of banned individuals or communities moved to alternative platforms

03

METRICS TO UNDERSTAND THE IMPACT OF DEPLATFORMING

We identify a series of outcomes or metrics that could be used to better assess the complexity of the impact of removing Trump from major social media platforms. We’ve grouped these 13 outcomes into 5 categories.

Some of the metrics we propose will be easier to study (e.g. amount of problematic content) than others (change in radicalized beliefs), but all will involve substantial challenges. Attributing any observed change in outcomes to Trump’s deplatforming will be extraordinarily difficult. Importantly, our focus here is on providing a set of criteria that could guide our assessments of impact, rather than offering a plan for that assessment.

A. Content

The amount and prominence of hate speech, misinformation, and/or illegal content has been a major concern both of academic analysis of platforms (e.g. Grinberg et al., 2018)¹¹ and of industry transparency reports.¹² As discussed above, it also has anchored many of the existing efforts to assess the impact of Trump’s removal.

We offer four specific metrics related to content that can provide insight into the impact of removing Trump from major platforms. For each of the four, it is important that analysis considers the impact of the ban on content across *users*, *platforms*, and *media*. Analysis should consider not only how the bans impacted content from Trump, but also content produced and shared by other users. Analysis needs to examine both the impact of the ban on content on the platform that removed Trump, as well as the impact on other platforms that may have seen increases in Trump supporters (Sanderson et

7 Rauchfleisch, A., & Kaiser, J. (2021). Deplatforming the Far-right: An Analysis of YouTube and BitChute (SSRN Scholarly Paper No. 3867818). Social Science Research Network. <https://doi.org/10.2139/ssrn.3867818>.

8 Horta Ribeiro, M., Jhaver, S., Zannettou, S., Blackburn, J., Stringhini, G., De Cristofaro, E., & West, R. (2021). Do Platform Migrations Compromise Content Moderation? Evidence from r/The_Donald and r/Incels. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 316:1-316:24. <https://doi.org/10.1145/3476057>.

9 <https://www.washingtonpost.com/technology/2021/01/16/misinformation-trump-twitter/>.

10 Alba, D., Koeze, E., & Silver, J. (2021, June 7). What Happened When Trump Was Banned on Social Media. *The New York Times*. <https://www.nytimes.com/interactive/2021/06/07/technology/trump-social-media-ban.html>.

11 Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 US presidential election. *Science*, 363(6425), 374–378.

12 <https://transparency.fb.com/data/community-standards-enforcement/>.

al., 2021).¹³ Finally, given the interconnections between social media and other media types, analysis should examine how the ban impacted content on other media, including TV, radio, podcasts, and political ads (Benkler et al., 2018).¹⁴

While we have our own personal views on the question of whether decreased volume, distribution, and engagement of Trump's content is positive for our society, our intent here is to avoid those political judgments. Some people may view such decreases as evidence deplatforming worked and others might view them as evidence deplatforming failed. Those value-based determinations are not our focus. Independent of whether these impacts are seen as positive or negative, they are important metrics for assessing the impact of the intervention.

“We offer four specific metrics related to content that can provide insight into the impact of removing Trump from major platforms

1. Amount of Problematic Content

As described above, examining whether banning Trump increased or decreased the amount of problematic online content is one of the few metrics existing analysis employs (Dwoskin & Timberg, 2021).¹⁵ The landscape of such content has been seen as indicative of the type of environment that platforms provide to users.

Defining problematic content provides serious challenges. Different platforms prohibit or action different types of content. Misinformation, hate speech, and illegal content, may all be broadly described as problematic, but are all different. It is possible that Trump's banning had different impacts on different types of (problematic) content.

2. Reach of Problematic Content

Beyond assessing the quantity of problematic content, analysis should examine how banning Trump impacted the reach of problematic content. The gross amount of problematic content on a platform is in some ways less relevant than the number of people who saw that content. For example, if Trump followers moved to alternative platforms, they may have continued to post objectionable content, yet that content may have reached far fewer users.

However, users view huge quantities of content on social media each day, little of which is meaningfully considered. The relationship between viewership and impact is complex, and deeply contextual.

3. Engagement with Problematic Content

Assessing how removing Trump did or did not impact the engagement with problematic content — including likes, shares, or comments — across platforms could provide a better sense of how many people actively considered that problematic content, especially if considered across platforms.

As with the amount and viewership of content, a great deal of scholarship has complicated the relationship between engagement with content and the impact of that content (Bennett & Iyengar, 2008).¹⁶ Most importantly, people may like, share, or comment on a piece of content for a number of different reasons (Marwick, 2018;¹⁷ Pennycook & Rand, 2021).¹⁸ Distinguishing between types of engagement may provide a more granular account of the impact of removing Trump.

4. Substance of Problematic Content

The three metrics discussed above lend themselves to quantitative analysis: studying how the bans impacted amounts of content, reach, or engagement. Yet, we also need a sense of how the bans impacted the substance of that content: the narratives, themes, and arguments made. At the same time, while it is essential that we understand

13 Sanderson, Z., Brown, M., Bonneau, R., Nagler, J., & Tucker, J. (2021). Twitter flagged Donald Trump's tweets with election misinformation: They continued to spread both on and off the platform | HKS Misinformation Review. *Misinformation Review*, 2(4). <https://misinformation.hks.harvard.edu/article/twitter-flagged-donald-trumps-tweets-with-election-misinformation-they-continued-to-spread-both-on-and-off-the-platform/>.

14 Yochai Benkler, Robert Faris & Hal Roberts, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*, Oxford University Press, October 2018.

15 Dwoskin, E., & Timberg, C. (2021, January 16). Misinformation dropped dramatically the week after Twitter banned Trump and some allies. *Washington Post*. <https://www.washingtonpost.com/technology/2021/01/16/misinformation-trump-twitter/>.

16 Bennett, W. L., & Iyengar, S. (2008). A New Era of Minimal Effects? The Changing Foundations of Political Communication. *Journal of Communication*, 58(4), 707–731. <https://doi.org/10.1111/j.1460-2466.2008.00410.x>.

17 Marwick, A. (2018). Why Do People Share Fake News? A Sociotechnical Model of Media Effects. *Georgetown Law Technology Review*. <https://www.georgetownlawtechreview.org/why-do-people-share-fake-news-a-sociotechnical-model-of-media-effects/GLTR-07-2018/>.

18 Pennycook, G., & Rand, D. G. (2021). The Psychology of Fake News. *Trends in Cognitive Sciences*, 25(5), 388–402. <https://doi.org/10.1016/j.tics.2021.02.007>.

how the bans may have impacted problematic content, we should also understand how it has impacted acceptable content as well.

B. Networks

Beyond its potential impact on social media content, Trump's removal may have affected Trump's networks of influence and support on and off social platforms.

1. Distribution of Followers and Content

One analysis found that while the bans decreased Trump's reach on major platforms, his supporters on the mainstream platforms increased their activity to help share and distribute Trump's (off-platform) statements (Alba et al., 2021).¹⁹ Further analysis could better examine the impact of these shifts in the distribution patterns of Trump's content.

For example: irrespective of its impact on the total amount of views/engagements with content, by leading to an increase in the number of people willing to directly support Trump on platforms, the bans may have mediated Trump's influence. Recently, social movement scholars have emphasized the importance of networked organizing to supplement "easy" online activism (Tufekci, 2017).²⁰ We need to better understand if and how the bans resulted in a larger or more diverse network willing to support Trump and share content, and what impact this might have had on Trump's influence and on public discussion.

2. Cross Media influence

Many have observed that Trump has benefited from "free" media coverage, as news outlets across the political spectrum follow him closely (Wells et al., 2016;²¹ Lawrence & Boydston, 2016).²² Notably, journalists regularly covered Trump's posts on Twitter and Facebook, granting Trump significant influence over the news agenda.²³ It is important that we understand better how being removed from major platforms influenced both Trump's coverage in mainstream

news as well as his ability to shape the topics discussed across outlets.

“However, users view huge quantities of content on social media each day, little of which is meaningfully considered

C. Beliefs

While studying the impact of Trump's removal on content may help us understand how the ban might have impacted the social media landscape, we should also consider how that content may or may not have affected users. First, we consider measures that speak to the impact on users' beliefs.

For nearly a century, media effects research has complicated the relationship between viewing content and being impacted by that content (Bennett & Iengar, 2008). While media content can shape the issues or topics we care about (McCombs & Shaw, 1972),²⁴ impacts on opinions or actions are much harder to tease out. This does not, however, mean that (lack of) access to Trump's content had no impact on followers—only that we need research that can identify and detangle the complex impact from the constellation of forces shaping users' beliefs.

1. Radicalization or Extreme Partisanship

How did banning Trump impact the number of users holding extremely partisan or radical beliefs or opinions? While radicalization has taken on many different meanings, here we follow Marwick et al (2022),²⁵ defining radicalization as "the process whereby individuals come to adopt an 'extremist' mindset or, more directly, escalate from nonviolent to violent political action over time." Importantly, there is

19 Alba, D., Koeze, E., & Silver, J. (2021, June 7). What Happened When Trump Was Banned on Social Media. The New York Times. <https://www.nytimes.com/interactive/2021/06/07/technology/trump-social-media-ban.html>.

20 Tufekci, Z. (2017). *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. Yale University Press. <http://gen.lib.rus.ec/book/index.php?md5=b7f5b30b96ae4b3de38fe32ccfa5ac8b>.

21 Wells, C., Shah, D. V., Pevehouse, J. C., Yang, J., Pelled, A., Boehm, F., Lukito, J., Ghosh, S., & Schmidt, J. L. (2016). How Trump Drove Coverage to the Nomination: Hybrid Media Campaigning. *Political Communication*, 33(4), 669–676. <https://doi.org/10.1080/10584609.2016.1224416>.

22 Lawrence, R. G., & Boydston, A. E. (2017). What We Should Really Be Asking About Media Attention to Trump. *Political Communication*, 34(1), 150–153. <https://doi.org/10.1080/10584609.2016.1262700>.

23 <https://www.businessinsider.com/elon-musk-twitter-donald-trump-ban-amplified-right-wing-experts-2022-5?r=US&IR=T>.

24 McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, 36(2), 176–187.

25 Marwick, A., Clancy, B., & Furl, K. (2022). Far-Right Online Radicalization: A Review of the Literature. *The Bulletin of Technology & Public Life*. <https://citap.pubpub.org/pub/jq7l6jny/release/1>.

little evidence that online content “causes” radicalization in any direct sense (Marwick et al., 2022). Research on online radicalism has consistently found a weak relationship between viewing online content and becoming radicalized (Gil et al., 2015)—yet online platforms can play catalyzing roles in radicalization, including by normalizing extreme content (Munn, 2019)²⁶ and by aiding community formation and identity development (Markwick et al., 2022).

2. Substance of Extreme Views

It is important that analysis not only captures the change in intensity of beliefs, but also the *qualitative* difference in the content of radical beliefs. How did it impact the narratives circulating in radical communities across platforms?

3. Trust in or Views of Social Media

Analysis should also consider how the ban impacted users' trust in social media platforms. It could examine how a change in trust has impacted platform use and how it shapes and is shaped by a broader decline in trust across institutions.

Scholars and opinion polls have traced a notable decline in trust in nearly all institutions across several decades (e.g. Gallup, 2022);²⁷ today, major platforms see low levels of public trust (Kelly & Guskin, 2021).²⁸ Within this context, we need to better understand what impact the bans may have had on the broader rejection of platforms.

“*Scholars and opinion polls have traced a notable decline in trust in nearly all institutions across several decades*”

D. Actions

It is also important that we understand how Trump's removal impacted the actions of followers on and off social media. While data about online or offline violence could be useful,

qualitative analysis could help us better understand how followers understand and narrate the impact of Trump's removal.

1. Online Violence

Has Trump's removal impacted the number of examples of online violence, including harassment, stalking, or doxing? Has it impacted the forms or severity of online harassment?

2. Offline Violence

How has Trump's removal impacted both the frequency and the nature of offline violence, including hate crimes or politically motivated violence? While attributing observed differences in amounts or types of hate crimes to the ban on Trump is unlikely, we need to understand better how banning Trump, in conjunction with changes in other forces that can radicalize users together may have played a role in increasing or decreasing offline violence.²⁹ Interviews with perpetrators of such crimes could help us better understand the broad constellations of forces that combined to facilitate violence, and what role, if any, Trump may have played in it.

E. Politics

Finally, analysis should consider if removing Trump from the major platforms had explicit political impact. Without some account of the influence of deplatforming on Trump's political power and on the greater landscape of politics, we will be missing an important part of the story.

1. Trump's Political Influence

While it is difficult to operationalize political impact, analysis could explore if the bans impacted Trump's ability to raise money for himself and allies, the power of his endorsements, or his ability to influence the Republican party platform.

2. Likelihood of Increased Regulation of the Tech Sector

Congress, state governments, and governments in other countries are considering a wide array of reform proposals. In the wake of the decision to deplatform Trump, a number of government officials – not only Republicans in the United

26 Munn, L. (2019). Alt-right pipeline: Individual journeys to extremism online. *First Monday*. Available online at <https://firstmonday.org/ojs/index.php/fm/article/view/10108>.

27 <https://news.gallup.com/poll/1597/confidence-institutions.aspx>.

28 Kelly, H., & Guskin, E. (2021, December 22). Americans widely distrust Facebook, TikTok and Instagram with their data, poll finds. *Washington Post*. <https://www.washingtonpost.com/technology/2021/12/22/tech-trust-survey/>.

29 As a result of the 1990 Hate Crime Statistics Act, the Justice Department must collect and report annual data on hate crimes. Unfortunately, the latest data released is from 2019.

States,³⁰ but also leaders³¹ of other countries – pointed to the decision as evidence of platform power and the need for stronger government regulation.

Deplatforming Trump could affect the likelihood that technology reform is passed. For instance, Republicans have historically been skeptical of increasing government intervention in private industry, but have been more inclined³² to support regulation of the tech sector because of the perception that it is biased against conservatives. A number of antitrust and content regulation proposals included both Republican and Democrat co-sponsors.³³ Deplatforming Trump therefore may impact the likelihood that regulation is passed, which could in turn influence the product experience for users and competitive dynamics in the industry.

04 CONCLUSION

Touching off 18 months of debate, the decision to ban the sitting President of the United States was an unprecedented move by online platforms. As politicians and commentators on the left and right argue over the merits of deplatforming Trump, discussion has been hamstrung by not having a shared understanding of how we might assess either the impacts or the effectiveness of the bans. In this paper we aim to start the conversation by offering a set of metrics that could anchor deeper and more granular analyses of the impact of Trump's ban. Using these metrics to empirically assess the bans could result in a deeper understanding of how the ban has impacted online content, networks, and politics.

Although we recommend researchers conduct empirical analysis to better understand the impacts of deplatforming, we recognize that no matter how exhaustive the data, this issue is unlikely to be resolved definitively. Conflicting value judgments and political perspectives means this issue will continue to be hotly contested. Nevertheless, judgments by policymakers and platforms should be informed by the full range of impacts deplatforming has had across public discussion and public life. Those assessments must begin with a shared understanding of what "impact" might mean... ■

“*Touching off 18 months of debate, the decision to ban the sitting President of the United States was an unprecedented move by online platforms*”

30 <https://www.latimes.com/politics/story/2021-05-06/how-big-tech-pushed-the-gop-into-the-corner-of-bernie-sanders>.

31 <https://www.reuters.com/article/usa-trump-germany-twitter/germany-has-reservations-about-trump-twitter-ban-merkel-spokesman-says-idUSL8N2JM4ES>.

32 <https://www.vox.com/recode/2019/10/29/20932064/senator-josh-hawley-tech-facebook-google-mark-zuckerberg-missouri>.

33 <https://slate.com/technology/2021/03/section-230-reform-legislative-tracker.html>.

CPI SUBSCRIPTIONS

CPI reaches more than **35,000 readers** in over **150 countries** every day. Our online library houses over **23,000 papers**, articles and interviews.

Visit competitionpolicyinternational.com today to see our available plans and join CPI's global community of antitrust experts.

