September 24, 2020

The Honorable Jan Schakowsky
Chairwoman
Subcommittee on Consumer
Protection & Commerce
Committee on Energy & Commerce
U.S. House of Representatives
Washington, DC 20515

The Honorable Cathy McMorris Rodgers
Ranking Member
Subcommittee on Consumer
Protection & Commerce
Committee on Energy & Commerce
U.S. House of Representatives
Washington, DC 20515

Dear Chairwoman Schakowsky, Ranking Member McMorris Rodgers, and Members of the Subcommittee:

My name is Will Duffield, I am a policy analyst with the Cato Institute's Center for Representative Government. I would like to thank the Subcommittee on Consumer Protection and Commerce of the Committee on Energy and Commerce for convening this hearing on Internet Radicalization, on September 24, 2020, and for providing the opportunity to express my views regarding this topic.

Today I want to make two broad points. First, internet-borne extremism is not primarily the fault of algorithms, but the natural result of open communicative tools that allow users to create non-geographic communities of interest. As a result, efforts to quash extreme speech will necessarily suppress innocent speech alongside it. Second, platform counter-extremism efforts should be cognizant of the off-platform effects of their decisions. Deplatforming extremists doesn't make them disappear and may drive them towards more extreme communities.

Algorithms Aside
Internet extremism is not primarily an algorithmic phenomenon. Instead, it is a function of the internet's capacity to host non-geographic communities of interest. The internet's connective power has been long celebrated, from the early days of Usenet to Tahir Square. However, as capacity for interconnection has been increasingly utilized by domestic extremists, Americans have begun to see cracks in the promise of an open internet. The problem is often understood as primarily algorithmic – social media feeds users increasingly extreme content in an effort to keep them engaged, slowly inducing a radical worldview. This tidy concern ignores both individual user agency and the broader effects of the internet.

Prior to the internet, and particularly social media, it was more difficult to form communities around niche interests. If you were a trainspotter, a fan of a foreign sports team, or obsessed with Nazi occultism, apart from a few periodicals and the occasional distant convention, you had little ability to join a community of the likeminded. Mainstream media was gatekept against radical thought on both the right and left – Huey Newton or David Duke might be occasionally interviewed by the mainstream press, but their adherents couldn't imbibe a 24-7 diet of Black Panther or KKK videos. On the internet, however, there is more content catering to every particular subcultural niche than anyone could hope to consume in a lifetime.

A recent paper by Penn State University Professor Kevin Munger and PhD Candidate Joseph Phillips titled "A Supply and Demand Framework for YouTube Politics," examines the effects of the internet's informational hyperabundance on political YouTube content. They find that the 'zombie bite' narrative of recommendation-driven informational contagion fails to account for the post-2017 falloff in viewership of alt-right YouTube content, despite increasingly supply, and note that a 5-step random walk algorithm beginning with 'alt-lite' content only reaches alt-right content one of every 1,700 trips. Most radicalization seems to occur via the standard model of persuasion, not algorithmic rabbit holes. Instead of the contagion model, Munger and Phillips propose a supply and demand framework for understanding the prevalence of extremist content on YouTube. They write:

> To date, journalistic and scholarly work has argued YouTube's recommendation algorithm has led viewers to extremist content, radicalizing them to further-right views. We argue instead that YouTube has affordances that make content creation easy for fringe political actors who tap into an existing base of disaffected individuals alienated from the mainstream, encouraging parasocial relationships that serve as a stand-ins for real sociality.[1]

They point to YouTube features that make it a valuable platform for extremist content creators – Discoverability via titling and tagging, channel-based community formation tools, and an international userbase – however, these features are useful to anyone who wants their videos to reach a wide audience. On the demand side, a well-documented decline in trust for government, mainstream media, and academia since the 1960s combined with other socioeconomic factors created an unmet demand for alternative sources of information.

Clearly this suggests a social media reflective of increasing radicalization, rather than one that impels extremism via algorithm. Or, to the extent that the internet drives radicalization, it does so by allowing speakers to serve previously unmet demands for radical content.

In many cases, existing criminal networks and activities have simply moved online where it can be archived and recorded. When social media makes existing criminal or extremist activity newly visible or archivable, it is often blamed for the pathologies it reveals. In a recent Cato Institute Policy Analysis, Georgia Institute of Technology Professor Milton Mueller describes how blame for malicious activity is assigned to platforms rather than users.

> The kinds of human activities that are coordinated through social media, good as well as bad, have always existed. However, these activities were not visible or accessible to the whole of society. As conversation, socialization, and commerce are aggregated into large-scale, public commercial platforms, they become highly visible to the public and generate storable, searchable records. Social media make human interactions hypertransparent and

---

[1] Kevin Munger and Joseph Phillips, "Right-Wing YouTube: A Supply and Demand Perspective", International Journal of Press/Politics, Forthcoming

displace the responsibility for societal acts from the perpetrators to the platform that makes them visible.[2]

Even as platforms render human activity newly visible, their sheer scale precludes easy governance, while the velocity of viral sharing makes it hard to avoid closing the door after the horse has bolted. Platform users upload billions of pieces of content, in hundreds of languages and dialects, with contextually specific meanings, every day. They cannot possibly understand the meaning of each and every message flowing across their service.

In response to this deluge of speech, platforms must choose between accepting more false positives (i.e., incidentally, or inadvertently removing more innocent speech), or accepting more extremist content. Even if a platform's identification of extremist content is 99.9% effective, it will produce a tremendous number of false positives in real terms. Twitter users post around 500 million tweets every day. If Twitter's moderation process were 99.9% percent accurate, it would still erroneously remove 500,000 tweets a day.

This is why commitments to free expression and efforts to suppress the spread of extremist thought are unavoidably in tension. At scale, limiting extremist expression will necessarily limit other sorts of expression too. In some instances, platform efforts to combat extremism may limit counter speech and counter-extremism research. Researchers and journalists studying extremist violence in the Middle East are frequently banned alongside the terrorist accounts they follow.

Algorithmic moderation is not a panacea. In response to COVID-19, YouTube sent most of its human moderators home, relying more heavily on algorithms. The number of videos removed from YouTube doubled compared to the preceding reporting period, from 5,700,000 removals between January and March, to 10,800,000 removals between April and June.[3] The platform's algorithm's didn't suddenly become more accurate. Over the same period, YouTube "saw both the number of appeals and the reinstatement rate double from the previous quarter. Notably, the number of videos reinstated on appeal increased from 25% of appeals in Q1 to 50% of appeals in Q2."[4] Expecting algorithmic moderation to remove more extremist content would mean accepting more false positives.

There is no way to easily remove more extremist content without removing anodyne content alongside it. Meaning can be subjective or difficult to discern, one man's extreme speech is often another's impassioned political engagement. Extremists often respond to moderation efforts with attempts to obscure their meanings, adopting elaborate codes and in-jokes that make identifying extremist content more difficult. In some cases, mainstream symbols are deliberately adopted by extremists in order to spur false positives and create an 'us or them' mentality among populations at risk for radicalization.

[2] Milton Mueller, Challenging the Social Media Moral Panic: Preserving Free Expression under Hypertransparency, Cato Institute Policy Analysis, July 2019, https://www.cato.org/publications/policy-analysis/challenging-social-media-moral-panic-preserving-free-expression-under

[3] Google Transparency Report, YouTube Community Guidelines Enforcement, https://transparencyreport.google.com/youtube-policy/removals

[4] YouTube, "Responsible policy enforcement during Covid-19", August 25, 2020, https://blog.youtube/inside-youtube/responsible-policy-enforcement-during-covid-19

Off-Platform Effects Matter
Platforms should also give more thought to the off-platform effects of their decisions. No one wants extremist content or networks on their platform. However, when banned from mainstream platforms, extremists do not simply disappear, they migrate elsewhere.

Platform moderation can effectively drive extremists underground, making their activities less visible. When extremists are forced off open or public platforms, they often move to private or encrypted environments. While this may impair their ability to recruit and spread their ideas, it also renders them more difficult to track or address. Contemporary right extremist violence is usually carried out by lone radicalized individuals. Before their crimes, they were most active in internet spaces out of the public eye. Per Adi Robertson[5] at The Verge:

> Radicalization doesn't just happen on Facebook and YouTube either. Many of the deadliest far-right killers were apparently incubated on small forums: Christchurch mosque killer Brenton Tarrant on 8chan; Oregon mass shooter Chris Harper-Mercer on 4chan; Tree of Life Synagogue killer Robert Bowers on Gab; and Norwegian terrorist Anders Breivik on white supremacist sites including Stormfront, a 23-year-old hate site credited with inspiring scores of murders.

Heavy handed moderation may drive some users down the radicalization funnel more quickly by pushing them towards platforms or fora where extremism is ignored by moderators and celebrated by users. In the wake of Facebook's recent crackdown on Boogaloo content, some displaced users turned to 4chan's /pol for comfort and community. No doubt, many banned communities on Facebook hosted extreme, or unhealthy, discourse. However, there it was nonetheless governed by Facebook's content rules. Likewise, as subreddits such as r/The_Donald and r/weekendgunnit have moved to the .win domain, they are no longer governed by reddit's community standards.

There is a very real danger that hearings such as this one spur companies to remove more speech, particularly political speech, than they would otherwise. Most of the content hosted in these spaces is entirely legal and protected by the First Amendment. While private firms have the right to govern their fora as they see fit, when they remove content in response to threatened regulation, government officials may gain an effective power to police speech that they are denied by our constitution.

As increased moderation by edge providers such as YouTube or Facebook pushes both extremist and incidentally adjacent communities towards alternative platforms, they turn speech and extremism suppression efforts toward the infrastructure that supports alternative platforms. Payment processors and web hosting providers are increasingly expected to deny service to purportedly extremist-friendly platforms. While this may be an effective means of denying speech-tools to bad actors, it risks amplifying the effects of mainstream platform decisions whiles leaving disfavored but legitimate speech without a home. While there is no easy fix for

---

[5] Adi Robertson, "Telling People to Delete Facebook Won't Fix the Internet," *The Verge*, 9/4/2020, https://www.theverge.com/2020/9/4/21419993/the-social-dilemma-jeff-orlowski-netflix-movie-review-social-media-algorithms

the spread of extremist ideas, they should ideally be combatted at the source, with solutions that address the conditions that give rise to extremism, rather than limits on Americans' access to contemporary communication tools.

Sincerely,

/s/

Will Duffield
Policy Analyst
Center for Representative Government
Cato Institute