DOWN THE ALT-RIGHT RABBIT HOLE: THE ROLE OF SOCIAL MEDIA PLATFORMS IN THE RADICALIZATION OF DOMESTIC RIGHT-WING TERRORISTS

PILAR MAE HOYE

INTRODUCTION

On May 14, 2022, an eighteen-year-old white man walked into a Tops supermarket in a historically Black neighborhood of Buffalo, New York and opened fire. The details of this tragedy—which left ten people dead, three others wounded, and a community in mourning—do not need to be repeated here. Instead, the focus must be on the role the internet played in this young man's self-radicalization and the motivation behind this act of domestic terrorism.[1] This shooting is one in an "epidemic of mass shootings often perpetrated by young men radicalized online by an ideology of hate,"[2] in which the internet, and social media in particular, plays an increasingly significant role.

By his own account, the Buffalo shooter's radicalization to becoming a white supremacist terrorist began when he watched a clip of the 2019 terrorist attack on a mosque in Christchurch, New Zealand.[3] His radicalization continued through further use of social media, most notably the platform 4chan, on which he engaged with virulent racist and antisemitic content.[4]

Like the Christchurch shooter, the Buffalo shooter livestreamed his attack.[5] Livestreaming has become a popular tool among terrorists, used to instantly publicize their crimes and increase the number of people terrorized

[1] The shooter was charged on the state level with one count of domestic terrorism in the first degree, ten counts of first-degree murder, ten counts of second-degree murder as a hate crime, three counts of attempted second-degree murder as a hate crime, and one count of second-degree criminal possession of a weapon. He pleaded guilty to fifteen counts, including the ten counts of first-degree murder and the first-degree domestic terrorism charge. He is the first person in the history of New York state to be convicted of domestic terrorism. Jonathan Franklin & Emily Olson, *The Buffalo Shooting Suspect Pleads Guilty to State Murder Charges*, NPR (Nov. 28, 2022), https://www.npr.org/2022/11/28/1138700312/ [https://perma.cc/QK63-7MZE].

[2] OFF. N. Y. STATE ATT'Y GEN. LETITIA JAMES, INVESTIGATIVE REP. ON THE ROLE OF ONLINE PLATFORMS IN THE TRAGIC MASS SHOOTING IN BUFFALO ON MAY 14, 2022 1 (Oct. 18, 2022), https://ag.ny.gov/sites/default/files/buffaloshooting-onlineplatformsreport.pdf [https://perma.cc/2LY4-U2C4].

[3] *Id.* at 3.

[4] *Id.*

[5] This is an increasingly popular choice among domestic terrorists and perpetrators of mass shootings. As one author noted in discussing the El Paso shooting, "[the El Paso shooter] livestreamed his massacre from a helmet cam in a way that made the shooting look almost exactly like a First Person Shooter video game. This was a conscious choice, as was his decision to pick a sound-track for the spree that would entertain and inspire his viewers." Robert Evans, *The El Paso Shooting and the Gamification of Terror*, BELLINGCAT (Aug. 4, 2019), https://www.bellingcat.com/news/americas/2019/08/04/the-el-paso-shooting-and-the-gamification-of-terror/ [https://perma.cc/GH9A-M3GY].

by the attack.[6] Though Twitch—the platform on which the Buffalo shooter livestreamed his attack—and Facebook—on which the Christchurch shooter livestreamed his—both took down the attackers' videos, the universal maxim that something uploaded to the internet cannot truly be removed has proven true. Facebook only removed the video of the Christchurch attack after the attack ended, and the Buffalo shooter credits livestream footage of this attack as playing a role in his radicalization. Twitch ended and removed the Buffalo shooter's livestream after only two minutes.[7] Even this comparatively short video was sufficient to accomplish the shooter's goals of terrorizing a community and inspiring future shooters as it spread online.[8] Livestreaming is one example of the way social media platforms not only help accomplish terrorists' goals, but also radicalize future terrorists.

This Note examines the role algorithm-based social media platforms play in the radicalization of lone-acting, right-wing, domestic terrorists. Though more fringe, user-directed social media platforms—such as 4chan and Gab—are home to some of the most extreme radicalizing content, mainstream, algorithm-based social media provides a platform for radicalizing content that is often promoted by the platforms' algorithms. Some examples include YouTube, Facebook, and X (formerly, Twitter), which pose a greater danger because of the nature of these platforms.

Algorithm-driven social media networks aim to keep the user on the platform as long as possible; to that end, each network has developed its own algorithm, which takes a user's past interaction history and presents the user with new content the user may enjoy, thus encouraging the user to spend more time on the platform.[9] As a consequence, the user is presented with information they may have never sought out or, at least, may have taken longer to find. By contrast, user-driven networks allow the user to entirely direct what content they see. Put colloquially, you do not accidentally become a terrorist on 4chan; you have to seek out the radicalizing content, whereas on YouTube, the radicalizing content could be put in front of you by an algorithm and accidentally radicalize you.[10]

To comply with the First Amendment, any new regulation that Congress imposes on social media companies for the purpose of curtailing social media's use for terrorist recruitment would have to be narrowly tailored and use the least restrictive means possible of achieving its aims. "The First

[6] OFF. N. Y. STATE ATT'Y GEN. LETITIA JAMES, *supra* note 2, at 3 ("Livestreaming has become a tool of mass shooters to instantaneously publicize their crimes, further terrorizing the public and the communities targeted by the shooter.").

[7] *Id.* ("Twitch, the platform used to livestream this atrocity, disabled the livestream within two minutes of the onset of violence, an improvement over Facebook's response to the livestream of the Christchurch attack, where the video was only removed after the attack ended. But two minutes is still too much. Even this relatively short video is enough for the horrific content to spread widely and to inspire future shooters.").

[8] *Id.*

[9] Kevin Roose, *The Making of a YouTube Radical*, N.Y. TIMES (Jun. 8, 2019), https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html [https://perma.cc/4GU9-KUA3] ("YouTube's algorithms were designed to 'increase the time people spend online, because it leads to more ads.'").

[10] *Id.* ("Over years of reporting on internet culture, I've heard countless versions of Mr. Cain's story: an aimless young man — usually white, frequently interested in video games — visits YouTube looking for direction or distraction and is seduced by a community of far-right creators. Some young men discover far-right videos by accident, while others seek them out. Some travel all the way to neo-Nazism, while others stop at milder forms of bigotry.").

Amendment has no categorical exemption for hate speech; . . . its creation and distribution cannot, constitutionally, be unlawful."[11] The First Amendment does, however, have a categorical exemption for incitement speech,[12] a category into which terrorist recruitment material likely falls. In offering potential solutions to the problem of social media's role in radicalizing domestic terrorists, this Note will discuss obstacles to new regulation including First Amendment concerns and Section 230 of the Communications Decency Act of 1996.

This Note will not explore the use of social media investigation and analysis by domestic law enforcement for national security or law enforcement purposes, nor the Fourth Amendment implications of such use. In discussing laws which could regulate social media platforms, this Note will not consider existing laws which criminalize providing material support to foreign terrorist organizations.[13] Throughout this Note, the names of any of the terrorists discussed will not be provided. As noted by former Prime Minister of New Zealand, Jacinda Ardern, after the Christchurch shooting in 2019, one thing lone actors attempt to gain through these terrorist acts is notoriety, and attaching their names to their terrorist acts would further this agenda.[14]

I. BACKGROUND

A. WHAT IS DOMESTIC TERRORISM?

Under federal law, domestic terrorism is defined as:

> activities that involve acts dangerous to human life that are a violation of the criminal laws of the United States or any State; appear to be intended to intimidate or coerce a civilian population, to influence the policy of a government by intimidation or coercion, or to affect the conduct of a government by mass destruction, assassination, or kidnapping,

which occur primarily within the United States.[15] Incidents of domestic terrorism have been increasing since 2014.[16]

The *Biden National Strategy for Countering Domestic Terrorism* emphasizes that the definition of domestic terrorism should include any

---

[11] OFF. N. Y. STATE ATT'Y GEN. LETITIA JAMES, *supra* note 2, at 3 ("The First Amendment has no categorical exemption for hate speech; most of the content the shooter viewed is rankly offensive, but its creation and distribution cannot, constitutionally, be unlawful").

[12] *See* Brandenburg v. Ohio, 395 U.S. 444, 453 (1969).

[13] *See generally* 18 U.S.C. § 2339B(a)(1) (it is also a crime to "knowingly provide[] material support or resources to a foreign terrorist organization."). This Note deals with acts of terrorism and speech that incites terrorist activity, not with whether such speech—and the platforms that host it—constitutes material support or resources for a foreign terrorist organization.

[14] *Christchurch Shooting: Ardern Vows Never to Say Gunman's Name*, BBC NEWS (Mar. 19, 2019), https://www.bbc.com/news/world-asia-47620630 [https://perma.cc/AUJ2-RP8T]; Meghan Keneally, *New Zealand Prime Minister Jacinda Ardern Vows to 'Never' Say Mosque Shooter's Name*, ABC NEWS (Mar. 19, 2019) https://abcnews.go.com/ABCNews/zealand-prime-minister-jacinda-ardern-vows-mosque-shooters/story?id=61781072 [https://perma.cc/TZA4-LG58].

[15] NAT'L SEC. COUNCIL, NAT'L STRATEGY FOR COUNTERING DOMESTIC TERRORISM 8 (Jun. 2021).

[16] Seth G. Jones, *The Evolution of Domestic Terrorism*, CTR. FOR STRATEGIC & INT'L STUD. (Feb. 17, 2022), https://www.csis.org/analysis/evolution-domestic-terrorism [https://perma.cc/4R5C-7WGW].

"individual or group who engages in violence" and anyone who "incites imminent violence."[17] This expanded definition seems to encourage prosecution of not only the individuals who commit acts of domestic terrorism, but also those who encourage acts of domestic terrorism without getting their hands dirty.

The particular events and conditions that motivate political violence are subjective.[18] While the current definition of terrorism does not identify any political views of those committing the acts of violence that constitute domestic terrorism, the intelligence community now recognizes the differing threats posed by actors with differing political motivations.[19] One key component of the current domestic terrorist threat comes from "racially or ethnically motivated violent extremists and networks whose racial, ethnic, or religious hatred leads them toward violence, as well as those whom they encourage to take violent action."[20] Currently, "white supremacists and other far-right wing extremists are the most significant domestic terrorism threat facing the United States."[21] Many recent domestic terrorist events in the United States can be linked to a white supremacist or other right-wing ideologies.[22]

One such ideology is the "great replacement" conspiracy theory, which motivated—among others—the El Paso[23] and Buffalo[24] attacks. The "great replacement" theory posits that the increasing non-white population and decreasing white population will lead to an eventual "replacement" of white people by people of color, to the detriment of society.[25] This conspiracy theory demonstrates sociologist Rory McVeigh's argument that "relatively politically advantaged groups (white, middle-class Christians, for example) are mobilized by a fear that their sociocultural status, political power, and institutional privilege are under threat of being given to other groups."[26] According to McVeigh, this fear motivates a myriad of types of collective

[17] NAT'L SEC. COUNCIL, *supra* note 15, at 9.

[18] Joan Donovan, Kaylee Fagan, & Frances Lee, *"President Trump is Calling Us to Fight": What the Court Documents Reveal About the Motivations Behind January 6 and Networked Incitement*, TECH. AND SOC. CHANGE PROJECT & HARV. KENNEDY SCHOOL SHORENSTEIN CTR. ON MEDIA, POLITICS AND PUB. POL'Y at 5 (Working Paper, July 18, 2022), https://mediamanipulation.org/sites/default/files/media-files/j6_motivations_working_paper.pdf [https://perma.cc/A2AS-TEFD] ("[T]he events and conditions that may serve as a 'trigger' for political violence are entirely subjective, and determined by movement members and leadership.").

[19] NAT'L SEC. COUNCIL, *supra* note 15, at 8; *See also* FED. BUREAU OF INVESTIGATION & DEP'T HOMELAND SEC., DOMESTIC TERRORISM: DEFINITIONS, TERMINOLOGY, AND METHODOLOGY (Nov. 2020) https://www.fbi.gov/file-repository/fbi-dhs-domestic-terrorism-definitions-terminology-methodology.pdf/view [https://perma.cc/S2SF-ECNR].

[20] NAT'L SEC. COUNCIL, *supra* note 15, at 9.

[21] S. 894, 116th Cong. § 2 (2019).

[22] Weiyi Cai, Troy Griggs, Jason Kao, Juliette Love & Joe Ward, *White Extremist Ideology Drives Many Shootings*, N.Y. TIMES (Aug. 4, 2019) https://www.nytimes.com/interactive/2019/08/04/us/white-extremist-active-shooter.html [https://perma.cc/L6KH-3SWR].

[23] John Eligon, *The El Paso Screed, and the Racist Doctrine Behind It*, N.Y. TIMES (Aug. 7, 2019), https://www.nytimes.com/2019/08/07/us/el-paso-shooting-racism.html [https://perma.cc/GWE8-VH3C];

Yasmeen Abutaleb, *What's Inside the Hate-Filled Manifesto Linked to the Alleged El Paso Shooter*, WASH. POST (Aug. 4, 2019), https://www.washingtonpost.com/politics/2019/08/04/whats-inside-hate-filled-manifesto-linked-el-paso-shooter/ [https://perma.cc/M3WR-YZ38]; Evans, *supra* note 5.

[24] PAYTON GENDRON, THE BUFFALO SHOOTER MANIFESTO (2022).

[25] Masood Farivar, *What is the Great Replacement Theory?*, VOICE AM. NEWS (Aug. 12, 2017), https://www.voanews.com/a/what-is-the-great-replacement-theory-/6578349.html [https://perma.cc/8H6J-DA89].

[26] Donovan, *supra* note 18, at 6.

action—up to and including forms of political violence, such as domestic terrorism—in a misguided attempt to preserve the dominant group's power and privilege, even at the expense of others.[27] Until the "great replacement" conspiracy theory gained popularity in the United States in recent years, most domestic mass shootings did not appear to be ideologically motivated. For example, the gunman behind the 2012 Aurora, Colorado shooting had no known extremist motives. While there is a long history of other forms of politically motivated domestic terrorism in this country,[28] it is undeniable that as this conspiracy gained popularity, it inspired many domestic terrorist attacks, as "[a]ggrieved white men . . . turned to mass murder in service of hatred against immigrants, Jews, and others they perceive as threats to the white race."[29] In particular, the El Paso shooting underscored "the global spread of white supremacist ideology in the age of social media," demonstrating the unique ability of social media to spread terrorist ideology, especially right-wing conspiracies, and radicalize domestic terrorists.[30]

B. LONE OFFENDERS

Today's domestic terrorists are often lone offenders who mobilize to violence with little or no clear organizational structure,[31] an evident departure from the meticulously-organized foreign terrorist networks, such as Al Qaeda, that made up much of the terrorist threat in the beginning of the 21st Century.[32] According to the FBI, lone offenders pose one of the most significant threats to the United States today, and these lone offenders are radicalized online.[33] For example, in his manifesto, the Buffalo shooter clearly states that he is "not a direct member of any organization or group," though he claims to support many.[34]

The intelligence community believes that violent extremists who are formally aligned with an organized militia group ("militia violent extremists" or "MVEs") present the most lethal domestic violent extremism threat. But the intelligence community has also assessed that "lone offenders . . . adhering to a diverse set of violent extremist ideologies are more likely to carry out violent attacks . . . than organizations . . . ."[35]

Lone offenders are uniquely dangerous because of their ability to more easily avoid detection and to mobilize to violence quickly.[36] They are difficult to detect because they often radicalize online, usually through social

---

[27] *Id.*

[28] Farivar, *supra* note 25; *see also* B. Hoffman, *Right-Wing Terrorism in the United States*, VIOLENCE, AGGRESSION AND TERRORISM 1987, at 1.

[29] Tim Arango, Nicholas Bogel-Burroughs & Katie Benner, *Minutes Before El Paso Killing, Hate-Filled Manifesto Appears Online*, N.Y. TIMES (Aug. 3, 2019), https://www.nytimes.com/2019/08/03/us/patrick-crusius-el-paso-shooter-manifesto.html [https://perma.cc/ZKN8-PCQ4].

[30] *Id.*

[31] NAT'L SEC. COUNCIL, *supra* note 15, at 9.

[32] *See* MARC SAGEMAN, UNDERSTANDING TERROR NETWORKS 137–174 (2004), https://www.jstor.org/stable/j.ctt3fhfxz.1 [https://perma.cc/HG7L-NJ25].

[33] NAT'L SEC. COUNCIL, *supra* note 15, at 9.

[34] GENDRON, *supra* note 24, at 4.

[35] NAT'L SEC. COUNCIL, *supra* note 15, at 10.

[36] *Id.* ("These individuals often . . . mobilize to violence quickly. . . . [They] are challenging to identify, investigate, and disrupt.").

media.[37] "[A]ttackers often radicalize independently by consuming violent extremist material online and mobilize without direction from a violent extremist organization, making detection and disruption difficult."[38] Social media has facilitated terrorist groups' ability to radicalize and recruit individuals who are receptive to terrorist recruitment material and, in particular, has enabled unprecedented access to individuals in the United States.[39] These lone offenders "will continue to pose significant detection and disruption challenges because of their capacity for independent radicalization to violence, ability to mobilize discreetly, and access to firearms."[40]

In a 2016 analysis of the individual radicalization of foreign fighters in the United States—U.S.-based individuals who travel abroad to join foreign terrorist organizations[41]—the National Consortium for the Study of Terrorism and Responses to Terrorism ("START") found that the average duration of radicalization has decreased in recent years. In 2002, the average radicalization duration for individuals was approximately sixteen months, whereas the average duration of radicalization for individuals in 2015 was just under ten months.[42] In this same period, the internet played an increasingly pivotal role in the radicalization of domestic fighters, contributing to the radicalization of 83% of individuals in 2015 (up from 37% of individuals in 2002).[43] Individuals who radicalized wholly or partly online "used the internet to view extremist materials, research conflicts, groups, and attack methods, and participated in online communities of like-minded individuals."[44] This report does not highlight any particular behavior profile that can clearly identify an individual likely to self-radicalize[45] and travel to join a foreign terrorist organization, as the individuals studied "exhibited a range of behaviors prior to traveling to conflict zones,"[46] the final step in their radicalization. This report illustrates the crucial role the internet plays in radicalizing terrorists today as well as the intelligence community's concern about the ability of lone offenders to easily avoid detection in the early stages of radicalization.

Despite its prevalent use by the intelligence community and others studying domestic terrorism, the term "lone actor" may be a misnomer. Though these terrorist acts are committed by individuals who appear to act independently, these terrorists' so-called "independent radicalization to violence" requires an existing network of like-minded people motivating

[37] FBI, *Terrorism* (last visited Aug. 25, 2023), https://fbi.gov/investigate/terrorism [https://perma.cc/BB2R-QP65] ("These individuals often radicalize online and mobilize to violence quickly.").

[38] NAT'L SEC. COUNCIL, *supra* note 15, at 10.

[39] FBI, *supra* note 37.

[40] NAT'L SEC. COUNCIL, *supra* note 15, at 11.

[41] Though this report does not discuss domestic extremists, its data is nonetheless useful as it indicates broader trends in contemporary terrorist radicalization.

[42] MICHAEL JENSEN, PATRICK JAMES & HERBERT TINSLEY, OVERVIEW: PROFILES OF INDIVIDUAL RADICALIZATION IN THE UNITED STATES — FOREIGN FIGHTERS (PIRUS—FF) (START National Consortium for the Study of Terrorism and Responses to Terrorism, 2016).

[43] *Id.*

[44] *Id.*

[45] For the purposes of this Note, to "self-radicalize" means to be the sole director of one's own radicalization as opposed to being guided to radicalization by a member of a terrorist organization, either online or in-person.

[46] Jensen, James, & Tinsley, *supra* note 42.

them to violence. As one author noted, "it is foolish, bordering on suicidal, to attribute attacks like the El Paso shooting or the Gilroy Garlic Festival[47] shooting to 'lone wolves.' Both shooters were radicalized in an ecosystem of right-wing terror that deliberately seeks to inspire such massacres."[48]

Believing that "lone actors" exist completely independently of existing terrorist networks leads to underestimating their potential for mass violence. While the intelligence community is careful to acknowledge the potential for lone actors to escape detection and radicalize quickly, media discussion of lone actors often fails to acknowledge this. The media serves a crucial role in helping reduce subsequent attacks. As noted by Robert Evans:

> "Until law enforcement, and the media, treat these shooters as part of a terrorist movement no less organized or deadly than ISIS or Al Qaeda, the violence will continue. There will be more killers, more gleeful celebration of body counts on 8chan, and more bloody attempts to be the last killer's 'high score.'"[49]

Recent examples of independently radicalized lone actors include the El Paso Walmart shooter,[50] who seems to have been influenced in part by the 2019 Christchurch shooter,[51] the Buffalo Tops supermarket shooter,[52] and the Uvalde Robb Elementary School shooter. Each of these lone actors seem to have been radicalized, at least in part, on social media.

II. THE ROLE OF SOCIAL MEDIA

It is clear that in recent years, social media networks have played a vital role in disseminating terrorist recruitment material and radicalizing domestic terrorists.[53] The Biden Administration's National Strategy for Countering Domestic Terrorism specifically references social media as one of the new technologies creating an increased, emerging threat to public safety.[54] In recent years, domestic terrorists have often been informally-aligned, lone actors who mobilize to violence with little or no clear direction.[55] These lone actors radicalize independently, and "often consume material deliberately

[47] In 2019, a 19-year-old lone gunman killed three people and wounded seventeen others before killing himself after a shootout with responding police officers at the Gilroy Garlic Festival in Gilroy, California. The gunman is believed to have acted alone. On the day of the shooting, the gunman posted on a newly made Instagram account complaining about the event and crowding caused by "hordes of mestizos and Silicon Valley white twats." *Gilroy Garlic Festival Shooting: Alleged Shooter Screamed Out "I'm Really Angry"*, CBS NEWS BAY AREA (Jul. 29, 2019), https://www.cbsnews.com/sanfrancisco/news/Gilroy-garlic-festival-mass-shooting-alleged-shooter-screamed-out-im-really-angry [https://perma.cc/ND6K-USDJ].

[48] Evans, *supra* note 5.

[49] *Id.*

[50] Arango, *supra* note 29.

[51] Evans, *supra* note 5.

[52] OFF. OF THE N.Y. STATE ATT'Y GEN. LETITIA JAMES, *supra* note 2, at 3; Aaron Katersky & Bill Hutchinson, *Buffalo Mass Shooting Suspect 'Radicalized' by Fringe Social Media: N.Y. Attorney General*, ABC NEWS (Oct. 18, 2022), https://abcnews.go.com/US/buffalo-mass-shooting-suspect-radicalized-fringe-social-media/story?id=91670651 [https://perma.cc/Y77F-HQBU]. Eligon, *supra* note 23.

[53] NAT'L SEC. COUNCIL, *supra* note 15, at 9.

[54] *Id.*

[55] *Id.*

disseminated to recruit individuals to causes that attempt to provide a sense of belonging and fulfillment."[56]

The intelligence community has assessed that domestic terrorists "exploit a variety of popular social media platforms . . . to recruit new adherents . . . and disseminate materials that contribute to radicalization and mobilization to violence."[57] The question then, as posed by Andrew Marantz in his 2019 book *Antisocial: Online Extremists, Techno-Utopians, and the Hijacking of the American Conversation* (*Antisocial*), is: "[i]f social media [isn't] a good product, why [is] it so successful?"[58]

"Modern terrorism relies heavily on the internet."[59] Advances in technology have improved terrorists' ability to plan and coordinate attacks and have increased the potential devastation of these attacks, making the modern terrorist more sophisticated than their predecessors.[60] The internet is "global and diffusive," in that it is "decentralized, inexpensive, innovative, and allows terrorists to remain anonymous and operate clandestinely."[61] The ability to reach individuals worldwide without a significant increase in cost is perhaps the biggest advantage that the internet gives to modern terrorists seeking to recruit new followers.

Social media is perfectly suited to promote terrorist agendas. Social media platforms are "known for their ability to bring like-minded people together, and terrorist organizations utilize these sites to recruit, fundraise, and spread terrorist propaganda."[62] These platforms are convenient and inexpensive, which allows individuals and organizations to "expand their global reach, amass support from other like-minded extremists, and capitalize on a larger network of diverse talents and skills."[63] Because social media platforms play an integral role in terrorist operations, efforts to curb the terrorist threats online impose onto social media companies an obligation to "take responsibility for the global implications of their platform."[64] Some experts advocate for platforms to demonstrate this responsibility in the form of active content monitoring and taking down extremist content—especially because platforms, as non-government entities, are not constrained by the First Amendment—while others contend that companies alone cannot be relied upon to curb terrorist speech online.[65] It is clear, however, that because social media platforms are integral to modern terrorism, "ISPs are indirectly contributing to terrorist causes."[66]

---

[56] *Id.*

[57] *Id.* at 11.

[58] ANDREW MARANTZ, ANTISOCIAL: ONLINE EXTREMISTS, TECHNO-UTOPIANS, AND THE HIJACKING OF THE AMERICAN CONVERSATION 76 (2019).

[59] Raphael Cohen-Almagor, *Symposium: The Role of Internet Intermediaries in Tackling Terrorism Online*, 86 FORDHAM L. REV. 425, 428 (2017).

[60] *Id.*

[61] Nicole Phe, *Social Media Terror: Intermediary Liability Under the Communications Decency Act*, 51 SUFFOLK U. L. REV. 99, 124 (2018).

[62] *Id.* at 100.

[63] *Id.*

[64] *Id.* at 101.

[65] *See* Michael Lavi, *Do Platforms Kill?*, 43 HARV. J. L. & PUB. POL'Y 477, 497-99; Raphael Cohen-Almagor, *Symposium: The Role of Internet Intermediaries in Tackling Terrorism Online*, 86 FORDHAM L. REV. 425, 430. *See generally* Marantz, *supra* note 58.

[66] *Id.* at 124.

A. RADICALIZATION OF FAR-RIGHT EXTREMISTS ON ALGORITHM-BASED SOCIAL NETWORKS

Research suggests that mainstream, algorithm-based social media is particularly efficient in recruiting like-minded individuals to participate in political engagement[67] through their social networks.[68] This is also demonstrably true of algorithm-based social media's effect on terrorist radicalization, both foreign and domestic. Even if lone actors go on to consume more extreme content on the fringe, user-directed social media platforms, mainstream algorithm-based social media platforms are a crucial first step in their online self-radicalization. The El Paso shooter's Twitter profile, for example, "left fallow since April 2017, suggests that at that time he projected the image of a relatively normal Trump-supporting Republican.[69] However, mainstream algorithm-based social media platforms likely catalyzed his continued radicalization. For example, "Twitter's [algorithm] has been described by one analyst of violent online jihadism as providing 'robust tools…to aspiring extremists' and 'a running start for users who are interested in pursuing ideologically motivated violence."[70]

Though skeptics point to a long history of political polarization and filter bubbles in traditional media landscapes, social media has amplified this issue. Social media enables groups and networks to pull new members in and motivate them to action on a much larger scale than previously seen.[71] A 2022 analysis was conducted by Lyn Van Swol, Sangwon Lee, and Rachel Hutchins of participation in political protest and the Capitol insurrection on January 6, though its findings are applicable to right-wing domestic terrorism more broadly and more violently than just these events. According to this analysis, the conditions that lead to polarization include "homogeneity of opinions among members, social comparison to more extreme members, easy exit for dissenters, and social categorization of like-minded members and identification."[72] The ease of exit for dissenters is especially important to the creation of online ideological echo chambers because it leads to isolation from dissenting voices.[73] This resulting lack of dissent strips potentially ambivalent group members of an alternative to the potentially radicalizing rhetoric being shared by the group.[74]

Van Swol et al.'s research on social media's facilitation of protest participation is equally applicable to the radicalization of terrorists on

---

[67] *See* Lyn Van Swol, Sangwon Lee & Rachel Hutchins, *The Banality of Extremism: The Role of Group Dynamics and Communication of Norms in Polarization on January 6*, 26 GRP DYNAMICS: THEORY, RSCH. & PRAC. 239, 240 (2022).

[68] Here meaning an actual web of social connections, not just social media networks, though social media networks are a tangible manifestation of these connections.

[69] Evans, *supra* note 5.

[70] Derek O'Callaghan, Derek Greene, Maura Conway, Joe Carthy, & Pádraig Cunningham, *Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems*, 33 SOC. SCI. COMPUT. REV. 459, 474 (2015).

[71] Van Swol, Lee, & Hutchins, *supra* note 67.

[72] *Id.*

[73] *Id.* ("If those who disagree with the trajectory of a group can easily leave, this reduces the chance of pushback and dissent.").

[74] *Id.* at 246.

algorithm-based social media platforms, including right-wing domestic terrorists. The elements of social media's role in this process are:

> (a) consistently exposing users to like-minded political information/news (both through active search and incidental exposure through algorithms) that would breed and strengthen their negative attitude toward the status-quo through exposure to more arguments; (b) maximizing one's network's effect, where ideologically like-minded individuals can not only share protest-related information and strategies but also emotions, concerns, and grievances . . . ; and (c) producing a need for approval and belonging to and social comparison with like-minded ingroup members that may push a participant toward more extremity and even action.[75]

These factors, especially on algorithm-based platforms, create a uniquely dangerous opportunity for users to self-radicalize online.

1. The Dangers of Algorithms

Algorithms create a particular danger in the radicalization of right-wing domestic terrorists on social media platforms. Each platform has a different proprietary algorithm guiding the user experience. Users, for their part, are often unaware of how these algorithms shape their experience on social media platforms including how these algorithms can control a user's exposure to news and extremist content, such as terrorist recruitment material.[76] While it is true that, even in traditional media environments, users "engage in selective exposure and consume information in line with their political beliefs,"[77] algorithms amplify this by creating few opportunities for users to be exposed to opposite or alternate beliefs.[78] This becomes particularly dangerous in the context of conspiracy theory content or terrorist recruitment material.

Algorithms exist primarily to keep users on a platform for as long as possible to help the platform earn money from advertisers. Social media platforms provide an experience that is free to users while sustaining themselves by monetizing user attention through advertising, which requires prolonged user engagement to maximize their profits. As a result, platforms have designed algorithms that "increase engagement, often with false, inflammatory or tribalizing content that research shows travels much more easily on social media."[79] This is largely because research has shown that the content evoking intense emotions in users is most likely to catch and keep their attention. These emotions can span from curiosity to humor, lust, nostalgia, envy, and outrage, among others.[80] As a result, social media often "elevates the worst, most divisive content, paired with the 'aaaw' and

[75] *Id.* at 242.
[76] *Id.* at 246.
[77] *Id.* at 240.
[78] *Id.* at 242.
[79] Zeynep Tufekci, *We Pay an Ugly Cost for Ads on Twitter*, N.Y. TIMES (Nov. 4, 2022), https://www.nytimes.com/2022/11/04/opinion/elon-musk-twitter-free.html [https://perma.cc/B3PQ-967U].
[80] MARANTZ, *supra* note 58, at 79 (2019) ("Curiosity is not the only way to get clicks, of course. Humor also works, as do lust, and nostalgia, and envy, and outrage.").

affirming-type content that promotes in-group bonding."[81] Social media companies with their current financial structure have no incentive to steer users away from extreme content. As Tristan Harris, a former design ethicist at Google (YouTube's parent company), has phrased it: "If I'm YouTube and I want you to watch more, I'm always going to steer you toward Crazytown."[82]

Core design features of social media platforms are exploited to promote extremism.[83] Though each social media platform has its own proprietary algorithm, most social media algorithms share a common purpose: to keep the user engaged and on the platform for as long as possible. In furtherance of this goal, algorithms promote the most engaging content, which is often content that inspires an intense emotional response in the user. "The more incendiary the material, the more it keeps users engaged, the more it is boosted by the algorithm."[84] As a result, social media algorithms often promote more extreme content. And as users engage with the extreme content presented by the platform's algorithm, the algorithm continuously presents them with other similar content, increasing their exposure. The process does not require the users to actively search such content because it is readily presented to them. "Research shows that Facebook even directs users who 'like' one militia page toward other militia groups." Algorithms will even push such content to the users who do not follow any of the accounts carrying it merely because the content is popular, hoping to attract the user's attention.[85] Companies are likely aware of the ways their platforms are exploited to promote extremism[86] but, driven by profit which requires user engagement, choose not to change the algorithms creating this risk.

B. YOUTUBE CASE STUDY

Until recently, YouTube has largely been ignored in conversations about terrorist recruitment material online.[87] Most research on the role of social media in radicalization has focused on Facebook and Twitter as legacy, news-oriented social media sites.[88] However, "[g]iven its billion or so users, YouTube may be one of the most powerful radicalizing instruments of the 21st century,"[89] With an estimated 500 hours of content uploaded to the platform every minute, there is ample content for users to explore.[90] While it

[81] Tufekci, *supra* note 79.

[82] Roose, *supra* note 9.

[83] Ambassador Karen Kornbluh, *Disinformation, Radicalization, and Algorithmic Amplification: What Steps Can Congress Take?*, JUST SEC. (Feb. 7, 2022), https://www.justsecurity.org/79995/disinformation-radicalization-and-algorithmic-amplification-what-steps-can-congress-take/ [https://perma.cc/FLU6-K9P9].

[84] *Id.*

[85] *Id.* ("Algorithms designed to keep users online then boost the material into feeds of users like 'Carol' because it is popular—whether or not she chose to follow any of the accounts carrying it.").

[86] *Id.* ("The mechanics of our platform are not neutral.").

[87] Roose, *supra* note 9 (quoting Becca Lewis) ("YouTube has been able to fly under the radar because until recently, no one thought of it as a place where radicalization is happening.").

[88] Homa Hosseinmardi, Amir Ghasemian, Aaron Clauset, Markus Mobius, David M. Rothschild, & Duncan J. Watts, *Examining the Consumption of Radical Content on YouTube*, 118 PNAS 1, 1 (2021).

[89] Zeynep Tufekci, *YouTube, the Great Radicalizer*, N.Y. TIMES (Mar. 10, 2018), https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html [https://perma.cc/5BU5-5LWK].

[90] Roose, *supra* note 9.

is misleading to claim that YouTube is the sole platform on which users can self-radicalize, or that YouTube entirely drives radicalization with little to no user input,[91] YouTube is an excellent case study on the role algorithm-based social media plays in the radicalization of right-wing domestic terrorists. At the very least, YouTube is "part of a larger information ecosystem in which conspiracy theories, misinformation, and hyperpartisan content are widely available, easily discovered, and actively sought out."[92] Furthermore, YouTube is not only where young people get their information and entertainment, but also where creators broadcast overtly White supremacist political content and other potentially radicalizing content.[93] Though these creators are active on other algorithm-based social media platforms as well, "YouTube was their headquarters."[94]

YouTube's recommendation algorithm has taken many forms. Currently, as users watch a video there is a "recommended-videos sidebar"[95] with algorithm-selected videos suggested as the next video for a user to watch, sorted in order from most to least likely to grab the user's attention. "To populate the recommended-videos sidebar, [the algorithm] first compiles a shortlist of several hundred videos by finding ones that match the topic and other features of the one you are watching. Then, it ranks the list according to the user's preferences, which it learns by feeding all your clicks, likes, and other interactions into a machine-learning algorithm."[96] In order to keep a user on the platform—in this case, engaging with algorithm-suggested videos—the algorithm "tends to offer choices that reinforce what someone already likes or believes."[97] This can create an addictive echo chamber for the user, and "often rewards the most extreme and controversial videos, which studies have shown can quickly push people into deep rabbit holes of content and lead to political radicalization."[98]

The YouTube algorithm is responsible for 70% of what users watch on the video-sharing platform.[99] The algorithm has gone through many changes in YouTube's history—most notably, a 2012 shift from prioritizing "view counts" to prioritizing "total watch time."[100] But crucial to every incarnation of the platform's algorithm is a desire to keep the user on the platform as long as possible because the more time a user spends on YouTube, the more YouTube earns in ad revenue.[101] A recent change to YouTube's algorithm has

[91] Hosseinmardi, Ghasemian, Clauset, Mobius, Rothschild, & Watts, *supra* note 88.

[92] *Id.*

[93] Roose, *supra* note 9 (quoting Becca Lewis) ("But it's where young people are getting their information and entertainment, and it's a space where creators are broadcasting political content that, at times, is overtly white supremacist.").

[94] *Id.*

[95] Karen Hao, *YouTube is Experimenting With Ways to Make its Algorithm Even More Addictive*, MIT TECH. REV. (Sept. 27, 2019), https://www.technologyreview.com/2019/09/27/132829/youtube-algorithm-gets-more-addictive/ [https://perma.cc/PA4V-X2WD].

[96] *Id.*

[97] *Id.*

[98] *Id.; see* Roose, *supra* note 9.

[99] *Id.*; Joan E. Solsman, *YouTube's AI is the Puppet Master Over Most of What You Watch*, CNET (Jan. 10, 2018), https://www.cnet.com/tech/services-and-software/youtube-ces-2018-neal-mohan/ [https://perma.cc/3XPW-VENG].

[100] Roose, *supra* note 9.

[101] *Id.* ("Guillame Chaslot, a former YouTube engineer who has since become a critic of the company's recommendation system, said this year that YouTube's algorithms were designed to 'increase the time people spend online, because it leads to more ads.'").

been described as "a kind of long-term addiction machine,"[102] designed to keep users engaged and on the platform longer by guiding them to different parts of the platform rather than repeating recommendations of their existing interests.

Although the YouTube algorithm has no inherent preference for extreme political content, far-right content has proven some of the most successful content under this recommendation algorithm.[103] Given the sheer volume of content posted to YouTube, it is nearly impossible to individually review and remove each potentially harmful video. Coupled with an algorithm designed to keep users on the platform for as long as possible to maximize advertising dollars, YouTube has inadvertently created the perfect breeding ground for radicalization. The combination of "a business model that rewards provocative videos with exposure and advertising dollars, and an algorithm that guides users down personalized paths meant to keep them glued to their screens" has been shown to lead susceptible users down the path of radicalization laid out for them by the recommendation algorithm.[104] "YouTube leads viewers down a rabbit hole of extremism, while Google racks up the ad sales."[105]

This "rabbit hole of extremism"[106] has been shown to benefit far-right creators and ideas. YouTube "has become the single most important hub by which an extensive network of far-right influencers profit from broadcasting propaganda to young viewers."[107] Researchers have found that "YouTube's algorithms created an isolated far-right community . . . and promoted misinformation."[108] It has been reported that a user starting from factual videos about the flu vaccine can be pushed by the algorithm to anti-vaccine conspiracy videos.[109] As Kevin Roose's article detailing the radicalization of Caleb Cain demonstrates, once a user has fallen down the rabbit hole, it is extremely difficult to escape[110] because of the algorithm's radicalizing feedback loop.

If YouTube had any desire to change its algorithm to avoid these aforementioned risks and problems, it is clearly capable of making changes. The company already makes many small changes to the algorithm each year, including ones designed to reduce the spread of conspiracy theories on the platform. One such change was to introduce a version of the algorithm that is activated after major news events "to promote videos from 'authoritative sources' over conspiracy theories and partisan content."[111] At the individual user level, YouTube has made changes to its recommendations system—the algorithm-based recommendations shown to individual users based on their watch history compared to other users—including one change where "a

[102] *Id.*
[103] *Id.*; Tufekci, *supra* note 89.
[104] Roose, *supra* note 9.
[105] *Id.*
[106] *Id.*
[107] Rebecca Lewis, *Alternative Influence: Broadcasting the Reactionary Right on YouTube*, DATA & SOC'Y (Sept. 18, 2018), https://datasociety.net/library/alternative-influence/ [https://perma.cc/TQ49-PKPM].
[108] Hao, *supra* note 95.
[109] Hosseinmardi, Ghasemian, Clauset, Mobius, Rothschild, & Watts, *supra* note 88.
[110] Roose, *supra* note 9.
[111] *Id.*

person who had watched a series of conspiracy theory videos would be nudged toward videos from more authoritative news sources."[112] Finally, as for specific videos, YouTube has said that a change to its algorithm "to reduce the spread" of objectionable videos that are objectionable, but not enough to be deleted outright "ha[s] resulted in significantly less traffic to those videos."[113]

Researchers have suggested other possible changes to YouTube's algorithm. One proposed update aims to target implicit bias, or "the way recommendations themselves can affect user behavior, making it hard to decipher whether you clicked on a video because you liked it or because it was highly recommended."[114] As a result of this implicit bias, the algorithm can push users away from the videos they "actually want to watch"[115] and toward more extreme content. Researchers proposed solving this problem by factoring in a video's rank in the recommendation sidebar every time a user clicks on one. Videos near the top of the sidebar would be given less weight in the algorithm's subsequent recommendations than videos that the user had to scroll to find.[116] When researchers tested this change live on YouTube, they found it significantly increased user engagement,[117] indicating that YouTube could stop steering users toward increasingly extreme content without losing their profits based on heightened user engagement.

Clearly, YouTube can make changes at every level—platform-wide, individual viewers' recommendation algorithms, and specific videos—to reduce the spread of conspiracy theory content on its platform. Changes to the platform's algorithm that can beneficially reduce terrorist recruitment material online are clearly possible without destroying YouTube, so why not implement them across the board?

III. REGULATION

There are many potential solutions to the problem of social media's role in the radicalization of right-wing domestic terrorists, with some currently being considered by the Supreme Court this term or being debated in Congress. Given the prevalence of social media use today, the ever-increasing social media user base within the United States, and the proven impact of social media networks on domestic terrorist radicalization, Congress should impose some form of regulation on social media companies to prevent the use of their networks as terrorist recruitment tools. The current approach of social media platforms—an "after the fact, whack-a-mole approach"[118]—to content moderation is clearly insufficient. However, many have concerns about the potential side effects of commonly suggested regulations, such as reform to Section 230 of the Communications Decency Act ("Section 230").

[112] *Id.*
[113] *Id.*
[114] Hao, *supra* note 95.
[115] *Id.*
[116] *Id.*
[117] *Id.*
[118] Kornbluh, *supra* note 83.

The Supreme Court will likely intervene, eventually, regarding both the extent to which the government—at the state or federal level—can regulate social media companies and Section 230. The current circuit split between the Fifth and Eleventh Circuits over anti-content moderation laws passed recently in Texas and Florida, respectively, makes intervention by the Supreme Court likely in the near future.[119] This past term, the Supreme Court had the chance to interpret how Section 230 applies to algorithm-based social media platforms in *Gonzalez v. Google LLC*, 598 U.S. 617 (2023).[120] Plaintiffs in *Gonzalez* sought to impose liability onto social media platforms for their role in allegedly radicalizing terrorists who went on to commit the November 2015 Paris attacks. The Court declined to consider the question Plaintiffs posed about immunity under Section 230, instead relying on their recent decision in *Twitter, Inc. v. Taamneh*, 598 U.S. 471 (2023)[121] to hold that Plaintiffs "state[d] little if any claim for relief."[122]

While allowing social media companies to attempt to self-regulate has advantages, it is insufficient as a solution to radicalization and can have the opposite of its intended effect. Social media platforms, as private actors, can limit speech in ways the federal government cannot. However, their current business revenue models disincentivize imposing such limitations. Platform executives often stress the value of the Honorable Oliver Wendell Holmes, Jr.'s "marketplace of ideas,"[123] arguing that the best way to combat bad speech is with more speech. For example, Facebook executive Simon Milner has "stressed the liberal concept of fighting opinions with opinions and argued that Facebook's officers are not equipped with the ability and knowledge to identify 'bad speech' as distinct from 'good speech.'"[124] Social media platforms always aim to encourage users to spend more time on the platform,[125] guided by the need to preserve advertising-driven revenue.[126]

[119] NetChoice, LLC. v. Paxton, 49 F.4th 439, 439 (5th Cir. 2022); NetChoice, LLC. v. Att'y Gen., Fla., 34 F.4th 1196, 1196 (11th Cir. 2022).

[120] The plaintiffs in *Gonzalez* were relatives of an American citizen who was killed in the 2015 Paris ISIS attacks. Plaintiffs sued defendant Google LLC under the Antiterrorism Act of 1990 ("ATA"), alleging in their operative complaint that Google was liable under the ATA for providing resources and assistance to ISIS through Google's ownership of YouTube. ISIS—among other terrorist organizations—is known to use YouTube to disseminate terrorist recruitment material. Plaintiffs further alleged that, despite YouTube's policies prohibiting terrorist content on its platform, YouTube failed to block ISIS's use of the platform. Most importantly, Plaintiffs alleged that YouTube's recommendation algorithm "assist[ed] ISIS in spreading its message." Holding Google liable in any capacity, however, required the Supreme Court to hold that any of Google's alleged actions fall outside the scope of protection offered by Section 230, which lower courts declined to do. Brief for the United States as Amicus Curiae Supporting Vacatur, at 5–6, *Gonzalez v. Google, LLC*, 2 F.4th 871 (2023) (No. 21-1333).

[121] Plaintiffs in both *Gonzalez* and *Twitter* attempted to impose secondary liability on platforms via laws that do not apply to this Note. The ATA and Justice Against Sponsors of Terrorism Act ("JASTA") require that, for a victim to seek compensation under these laws, the victim must have been injured by an act of "international terrorism" committed by "a foreign terrorist organization designated as such as of the date on which such act of international terrorism was committed, planned, or authorized." Twitter, Inc. v. Taamneh, 598 U.S. 471, 495 (2023) (citing § 2333(a),(d)) (internal quotations omitted). This Note exclusively discusses domestic acts of terror committed by persons not affiliated with formal terrorist organizations. Therefore, these laws do not apply to this discussion.

[122] Gonzalez v. Google LLC, 598 U.S. 617, 622 (2023).

[123] Abrams v. United States, 250 U.S. 616 (1919) (Holmes, J., dissenting).

[124] Cohen-Almagor, *supra* note 59, at 431.

[125] MARANTZ, *supra* note 58, at 80.

[126] Roose, *supra* note 9 ("YouTube's algorithms were designed to 'increase the time people spend online, because it leads to more ads.'"). *See generally* Catherine Prince, *Trapped — The Secret Ways Social Media is Built to be Addictive (and What You Can Do to Fight Back)*, BBC SCIENCE FOCUS (Oct.

Therefore, platforms are unlikely to impose sufficient regulations to stop the use of these platforms for terrorist recruitment.[127] Research makes it clear that the social media companies' "current strategy of post hoc, individual take-downs is grossly insufficient to address this systemic vulnerability."[128] The need for some solution is clear, as "[m]orally speaking, we cannot be neutral regarding such alarming speech."[129]

A. REGULATION ROADBLOCKS

1. Legislative Limits: Section 230

Section 230 was instrumental to creating the modern internet because without this legislation, social media networks—and all other internet companies—would be subject to strict liability for every message and post made on their platforms. By shielding early internet companies from liability and encouraging attempts at good-faith moderation, "Congress enabled a range of innovative new websites to offer social networking, video sharing, and other 'Web 2.0' services that have transformed how we do business and socialize online,"[130] thereby allowing for the creation of the modern internet. However, Section 230 was not intended to "create a lawless no-man's-land on the Internet,"[131] nor to completely insulate platforms from liability for their role in promoting incitement speech and terrorist recruitment material online.

Section 230 protects only certain defendants from certain claims; specifically, it protects websites and other online platforms from claims seeking to treat them as the "publishers" or speakers of third-party information, rather than mere neutral conduits for the information.[132] Activities traditionally classified as "publishing" include "reviewing, editing, and deciding whether to publish or to withdraw from publication third-party content."[133] According to the Department of Justice, "'[p]ublisher' is best read in this context to refer to one who commits the common-law act of 'publication': the communication or dissemination of expressive material to another. Claims alleging liability based on a platform operator's failure to block or remove material created and posted by third parties meet this element, regardless of the precise cause of action."[134]

29, 2018) https://www.sciencefocus.com/future-technology/trapped-the-secret-ways-social-media-is-built-to-be-addictive-and-what-you-can-do-to-fight-back [https://perma.cc/G3KK-C5UG].

[127] OFF. OF THE N.Y. STATE ATT'Y GEN. LETITIA JAMES, *supra* note 2, at 3–5("In the absence of changes to the law, platforms like 4chan will not take meaningful action to prevent the proliferation of this kind of content on its site. . . . We can no longer rely entirely on the industry to regulate itself through voluntary commitments.").

[128] Kornbluh, *supra* note 83.

[129] Cohen-Almagor, *supra* note 59, at 429.

[130] *Id.* at 432.

[131] Phe, *supra* note 61 (citing Fair Hous. Council of San Fernando Valley v. Roomates.Com, LLC, 521 F.3d 1157, 1172 (9th Cir. 2008)).

[132] Gonzalez v. Google, LLC, 2 F.4th 871, 889 (9th Cir. 2021) ("§ 230 protects from liability only a specific class of defendants facing a particular type of claim — i.e., it protects providers and users of interactive computer services from claims seeking to treat them as publishers or speakers of information provided by others." (citing *Barnes* 570 F.3d 1096, 1100–01 (9th Cir. 2009))).

[133] Barnes v. Yahoo!, Inc., 570 F.3d 1096, 1102 (9th Cir. 2009).

[134] Brief for the United States as Amicus Curiae Supporting Vacatur, at 9, *Gonzalez*, 2 F.4th 871 (No. 21-1333).

Section 230(c)(2) emphasizes that platforms that engage in good faith moderation efforts retain the immunity provided, even though upon engaging in good faith moderation efforts under tort law, platforms would typically lose protection because they could no longer hide behind an ignorance defense.

The Ninth Circuit laid out the factors required for a publisher to receive Section 230 immunity in *Barnes v. Yahoo!, Inc.* Section 230(c)(1):[135] "(1) a provider or user of an interactive computer service (2) whom a plaintiff seeks to treat, under a state law cause of action, as a publisher or speaker (3) of information provided by another information content provider."[136] The law precludes a plaintiff's claim only if all three elements are met.[137] Speaking generally about claims against social media platforms, "the first prong is easily satisfied because social media websites clearly provide interactive computer services."[138] The substantial question for the court arises under the second prong. Under the second prong, the question is whether the social platform was in any way responsible for the "creation or development of the information" at issue in the case.[139] In answering that question, the court may also consider whether the platform has "materially contributed" to the third party's unlawful conduct.[140] Finally, under the third prong, the court must analyze the plaintiff's theory of liability to determine whether the cause of action falls within the purview of traditional editorial functions.[141] This analysis is required even if a plaintiff attempts to circumvent Section 230 by pursuing liability through another legal avenue, such as tort liability.[142]

Platforms are only immune from claims arising from hosting information posted by a third party. Platforms are not immune from liability for information "that they themselves create or develop, in whole or in part."[143] If the platform is "'responsible, in whole or in part' for 'creating or developing' the actionable material," the platform can be held liable because it then becomes the aforementioned third-party poster.[144]

Additionally, platforms do not have immunity if they provide advertisers with "tools designed to target ads to users based on sex, race, or other protected characteristics in areas covered by civil rights laws," or if the platform's own advertising delivery algorithms are allegedly inherently discriminatory.[145] Each example provided in this list comes from a court's decision to read Section 230 narrowly in a given case. For example, in *Fair*

[135] Barnes v. Yahoo!, Inc., 570 F.3d 1096, (9th Cir. 2009); "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider." 47 U.S.C. § 230(c)(1).

[136] *Barnes*, 570 F.3d at 1100–01.

[137] Brief for the United States as Amicus Curiae Supporting Vacatur, at 8, *Gonzalez*, 2 F.4th 871 (No. 21-1333).

[138] Phe, *supra* note 61, at 111.

[139] *Id.*

[140] *Id.*

[141] *Id.*

[142] Zeran v. Am. Online, Inc., 129 F.3d 327, 332 (1997) ("Because the publication of a statement is a necessary element in a defamation action, only one who publishes can be subject to this form of tort liability.")

[143] Brief of American Civil Liberties Union of Northern California & Daphne Keller as Amici Curiae in Support of Respondents, at 9, Gonzalez v. Google, LLC, 2 F.4th 871 (2023) (No. 21-1333) (citing Fair Hous. Council of San Fernando Valley v. Roomates.com, LLC, 521 F.3d 1157, 1172 (9th Cir. 2008)).

[144] *Id.* at 13.

[145] *Id.* at 11.

Housing Council of San Fernando Valley v. Roomates.com, the provider was held liable because it "materially contribut[ed] to [the content's] alleged unlawfulness."[146] This assessment is crucial to balancing the aims of Section 230 with other laws. The Department of Justice, in an amicus brief, argued that "[a]n overly broad reading of Section 230(c)(1) would undermine the enforcement of other important federal statutes by both private plaintiffs and federal agencies."[147]

However, Section 230 does not bar claims based on social media platforms' targeted recommendations of terrorist recruitment materials, such as videos featured in YouTube's "Up Next" algorithm.[148] Because holding platforms liable for algorithm recommendations turns on the platforms' own conduct and communications, algorithm recommendations are editorial actions taken by the social media platforms that fall outside the scope of protection offered by Section 230.

2. Regulating Speech & Constitutional Concerns

In considering potential government regulation of online terrorist speech, First Amendment concerns naturally arise. This is true even in the context of bold actions that private companies might undertake, as "concerns regarding the potential suppression of free speech are likely to arise if social media websites immediately remove suspected terrorist accounts."[149] According to Jameel Jaffer, executive director of the Knight First Amendment Institute at Columbia University, "[t]he First Amendment protects Americans' right to access social media platforms of their choice. . . . Banning or restricting access to social media is a hallmark of authoritarian regimes . . . ."[150] The Supreme Court is likely to attempt to resolve a current circuit split over social media companies' abilities to regulate speech on their platforms.[151] Until the Supreme Court does so, current law holds that social media companies have the ability to moderate speech on their platforms under Section 230, and the government has restrictions on its ability to regulate speech imposed by the First Amendment.

Under the First Amendment, the government can only impose content-based regulations on protected speech if the government proves that the regulation (1) furthers a compelling government interest and (2) is narrowly tailored, using the least-restrictive means of achieving its goals.[152] However, the government is free to impose regulations on unprotected forms of speech, such as incitement speech.

Since incited speech is unprotected, any government regulations of this form of speech need only meet the current incitement test under *Brandenburg v. Ohio*, 395 U.S. 444 (1969). Under *Brandenburg*, both state

[146] *Roomates.com*, 521 F.3d at 1167–68.

[147] Brief for the United States as Amicus Curiae Supporting Vacatur, at 2, *Gonzalez*, 2 F.4th 871 (No. 21-1333).

[148] *Id.* at 10.

[149] Phe, *supra* note 61, at 130.

[150] Press Statement, Knight Institute Comments on Proposed TikTok Ban, (March 22, 2023) (on file with Knight First Amend. Inst. at Columbia Univ.), https://knightcolumbia.org/content/knight-institute-comments-on-proposed-tiktok-ban [https://perma.cc/BZP5-BEC4].

[151] NetChoice, LLC v. Paxton, 49 F.4th 439 (2022); NetChoice, LLC v. Att'y Gen., Fla., 34 F.4th 1196 (2022).

[152] Reed v. Town of Gilbert, Ariz., 576 U.S. 155, 155 (2015).

and federal governments can regulate speech without violating the First Amendment if the speech incites violence, creating a "clear and present danger."[153] A "clear and present danger"[154] is defined as a likelihood of "imminent lawless action" that the speaker intends to cause.[155]

In analyzing whether certain forms of terrorist recruitment material are unprotected incitement speech, the crucial components of the *Brandenburg* test are (1) the imminence of the potential danger and (2) the intent of the speaker. The *Brandenburg* test is reflected in the Biden Administration's expanded definition of domestic terrorism.[156] The specification of "incites *imminent* violence" (emphasis added) ensures this definition reflects the current imminence test for incitement speech.[157] This definition, however, makes no mention of the intent of the speaker. Any regulation or legal action would have to establish a clear link between the information shared—for example, repeating the "great replacement" theory discussed above—and a desire by the speaker to inspire a listener to unlawful violence.

The *Brandenburg* test, in its attempt to distinguish inciting speech from "mere advocacy,"[158] leaves several crucial points ambiguous: How imminent must the intended lawless action be to warrant a restriction on speech? Does this intent include the doctrine of double effect, where one can be charged with intending both the desired effect of one's speech and the unknown, if not desired, effects of the same speech?[159] What is the relationship between the probability of harm occurring and the degree of harm caused under this test?

While the Biden administration's expanded definition of terrorism is not a legally binding restriction on speech and seems to comply with the *Brandenburg* test, any potential future legislation would need to comply with the *Brandenburg* test and would risk getting caught in one of the ambiguities identified above.

B. UNINTENDED CONSEQUENCES

It is, of course, entirely possible that the cure for this problem is worse than the disease, or at least would have far-reaching unintended consequences. A potential chilling effect on non-incitement speech is certainly a concern.[160] As is the fact that "states that regulate or influence platforms often also, intentionally or not, shape speech rules that the platforms apply in other countries,"[161] indicating potential international implications of any regulation imposed by the U.S. As noted by Derek O'Callaghan, Derek Greene, Maura Conway, Joe Carthy, and Pádraig

[153] *See Brandenburg*, 395 U.S. at 453 (Black, J., concurring) (citing Bridges v. California, 314 U.S. 252, 261 (1941)).
[154] *Id.*
[155] *Id.*
[156] NAT'L SEC. COUNCIL, *supra* note 15.
[157] *See Brandenburg*, 395 U.S. at 453 (Black, J., concurring); *see also* Holder v. Humanitarian L. Project, 561 U.S. 1, 43–44 (2010).
[158] Brandenburg v. Ohio, 395 U.S. 444, 448–49 (1969).
[159] *See* Abrams v. U.S. 250 U.S. 616, 626-27 (1919).
[160] Cohen-Almagor, *supra* note 59, at 431.
[161] Daphne Keller, *Who Do You Sue?*, *in* AEGIS PAPER SERIES 7 (THE HOOVER INST., 2019).

Cunningham, many commonly suggested interventions "raise the specter of social media companies policing political thought,"[162] which nobody wants.

Take, for example, the Fight Online Sex Trafficking Act (FOSTA) and the Stop Enabling Sex Traffickers Act (SESTA). The FOSTA-SESTA package clarified the United States' sex trafficking law, making it illegal to knowingly assist, facilitate, or support sex trafficking, and to amend Section 230's safe harbors to exclude enforcement of federal or state sex trafficking laws from its immunity. FOSTA-SESTA has been criticized by advocacy groups due to concerns about the potentially dangerous and disproportionate impact of this package on sex workers and sex educators online. Furthermore, free speech advocates—including the ACLU—argued that FOSTA-SESTA placed an unnecessary burden on internet companies to handle user-generated content, now that these companies no longer had any safe harbor provisions to protect their good-faith content moderation. In response to FOSTA-SESTA, several social media platforms enacted policy changes severely restricting the posting of sexual content on their platforms,[163] and some platforms closed entirely.[164] These policy changes have been shown to disproportionately impact LGBTQ+ content creators and sex educators. Meanwhile, according to a 2021 Government Accountability Office ("GAO") study, only one case was brought under FOSTA-SESTA in the three years since the package was passed.[165] FOSTA-SESTA has chilled protected speech, while failing to accomplish its goal of stopping sex trafficking online.[166] Could the same not also be true for any regulation targeting terrorist speech on social media platforms?

Considering regulation of algorithm-based social media platforms poses another problem: how can users navigate the firehose of information uploaded to social media platforms daily without some sort of filter? Opponents of imposing liability claim that platforms "must make decisions about how to organize and display . . . content if the site is to be usable."[167] "With two billion monthly active users uploading more than 500 hours of video every minute, YouTube's traffic is estimated to be the second highest of any website . . . 94 percent of Americans ages 18 to 24 use YouTube, a higher percentage than for any other online service."[168] Without YouTube's recommendation algorithm, how can users begin to navigate this library of

[162] O'Callaghan et. al., *supra* note 70, at 474.

[163] Paris Martineau, *Tumblr's Porn Ban Reveals Who Controls What We See Online*, WIRED (Dec. 4, 2018, 2:07 PM), https://www.wired.com/story/tumblrs-porn-ban-reveals-controls-we-see-online/ [https://perma.cc/LTK4-GQBP]; Alexander Cheves, *The Dangerous Trend of LGBTQ+ Censorship on the Internet*, OUT (Dec. 6, 2018, 12:16 PM), https://www.out.com/out-exclusives/2018/12/06/dangerous-trend-lgbtq-censorship-internet [https://perma.cc/F4W5-CVCX]; Elliot Harmon, *Facebook's Sexual Solicitation Policy is a Honeypot for Trolls*, ELEC. FRONTIER FOUND. (Dec. 7, 2018), https://www.eff.org/deeplinks/2018/12/facebooks-sexual-solicitation-policy-honeypot-trolls [perma.cc/NER3-2RV9].

[164] Samantha Cole, *Craigslist Just Nuked Its Personal Ads Section Because of a Sex-Trafficking Bill*, VICE: MOTHERBOARD (Mar. 23, 2018, 5:18 AM), https://www.vice.com/en/article/wj75ab/craigslist-personal-ads-sesta-fosta [https://perma.cc/DG3P-XE9S].

[165] *Sex Trafficking: Online Platforms and Federal Prosecutions*, U.S. GOV'T ACCOUNTABILITY OFF. (June 2021), https://www.gao.gov/assets/gao-21-385.pdf [https://perma.cc/UJ4U-H6Z5].

[166] Kendra Albert, Emily Armbruster, Elizabeth Brundige, Elizabeth Denning, Kimberly Kim, Lorelei Lee, Lindsey Ruff, Korica Simon & Yueyu Yang, *FOSTA in Legal Context*, 52 COLUM. HUM. RIGHTS L. REV. 1084 (2021).

[167] Brief of American Civil Liberties Union of Northern California & Daphne Keller as Amici Curiae in Support of Respondents, at 15, Gonzalez, 2 F.4th 871 (No. 21-1333).

[168] Roose, *supra* note 9.

video content? The ACLU raises this argument in an amicus brief in *Gonzalez v. Google*.[169] This argument, however, ignores that most social media platforms have been algorithm-less—and successful—before. For example, Instagram sorted posts chronologically until 2016, and recently revived this feature in the face of increasing public demand for the platform to move away from an algorithm-organized feed.[170]

IV. A POTENTIAL SOLUTION

Clearly, the status quo cannot continue. While in theory, hate speech, conspiracy theories, and other potentially radicalizing content can exist online without endangering the public, in reality, they pose a significant threat. As Matthew Feldman, Director of the Center of Analysis of the Radical Right, observed, "[a]llowing hate speech to fester is like leaving a wound unattended. At best it is unpleasant. At worst it can make other parts poorly or sick, and in extremis even kill."[171] The way algorithm-based social media networks are currently designed allows people to exploit these tools to promote extremism and radicalize new terrorists. Proposed solutions, however, are often both over and under-inclusive in their approach.[172] Despite the challenges of adapting existing legal principles to emerging technology, "the basic principles of freedom of speech and the press, like the First Amendment's command, do not vary when a new and different medium for communication appears."[173]

First, any proposed solution cannot be applied broadly to all social media networks for the same reason that this Note focuses on some networks and not others: there are inherent differences between types of social media networks that make it impossible to universally apply a one-size-fits-all solution. Furthermore, no one solution can completely solve this problem. The best answer is likely a combination of regulations and private litigation. This Note discusses one possible solution among many, in the context of recent Supreme Court cases and current findings about the impact of social media platforms in radicalization.

---

[169] Brief of American Civil Liberties Union of Northern California & Daphne Keller as Amici Curiae in Support of Respondents, at 23, *Gonzalez*, 2 F.4th 871 (No. 21-1333) ("If a platform will lose legal immunity for making recommendations of offending content, it may seek to eliminate ranking entirely, creating a site consisting of disorganized content that will satisfy no one.").

[170] Lindsay Lowe, *Instagram Brings Back Chronological Feeds — With a Catch*, TODAY (Mar. 24, 2022, 9:02 AM), https://www.today.com/news/news/instagram-chronological-feed-back-default-catch-rcna21356 [https://perma.cc/JJF2-JTCN]; Instagram (@Instagram), X (Mar. 23, 2022, 12:58 PM), https://twitter.com/instagram/status/1506722074810720257?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1506722074810720257%7Ctwgr%5E894abf8b9ee959160450e0ddba569629ba32fc29%7Ctwcon%5Es1_&ref_url=https%3A%2F%2Fiframe.nbcnews.com%2F1pOqBu5%3F_showcaption%3Dtrueapp%3D1 [https://perma.cc/P5LR-C6DX]; Adam Mosseri (@mosseri), INSTAGRAM (Mar. 23, 2022), https://www.instagram.com/reel/CbdD0MCJyZ0/?utm_source=ig_embed&ig_rid=3727ee31-dd64-4786-a849-eff54269deb6 [https://perma.cc/DJ6P-YR4M].

[171] David Gilbert, *Here's How Big Far-Right Social Network Gab Has Actually Gotten*, VICE: NEWS (Aug. 16, 2019, 11:35 AM), https://www.vice.com/en/article/pa7dwg/heres-how-big-far-right-social-network-gab-has-actually-gotten [https://perma.cc/GZV3-ZJ3R].

[172] Kornbluh, *supra* note 83 ("Design features of social media platforms are exploited to promote extremism. The platforms' after the fact, whack-a-mole approach to content moderation is insufficient. However, Section 230 reform—a popular rallying cry—is a blunt instrument that may lead to unintended suppression of important speech and not address radicalization.").

[173] Brown v. Ent. Merchs. Ass'n, 564 U.S. 786, 790 (2011).

In using algorithms to promote content, social media networks cease to be neutral hosts of information and begin to function like publishers; therefore, such activity should be exempted from the broad immunity provided by Section 230. This is similar to the theory of liability offered by the plaintiffs in *Gonzalez v. Google*. This is also like the civil liability for platforms advocated by New York State Attorney General Letitia James in the report on the role of online platforms in the radicalization of the Buffalo shooter.[174]

Imposing liability in this manner does not demand that all social media platforms become recommendation-free hellscapes[175] that are unpleasant—if not impossible—to use, as opponents would seem to think.[176] Nor does it result in the wholesale destruction of Section 230's immunity protections.[177] In fact, one of the intentions behind Section 230 was to "remove disincentives for the development and utilization of blocking and filtering technologies,"[178] indicating that the blocking and filtering required to change a platform's algorithms are already incorporated into the scope of Section 230. As discussed above, if a platform can be held liable based on its advertising algorithms, why can a platform not be held liable based on its recommendation algorithms? Imposing liability merely encourages platforms to be more careful with the design of their algorithms, perhaps returning to features users seem to prefer, such as Instagram's chronological feed.

CONCLUSION

In sum, the algorithms used by many popular social media platforms fall outside the scope of the protections provided by Section 230, as it is an editorial feature which makes these platforms publishers of recommended content. Platforms can and should be held liable for any terrorist action taken based on radicalization spurred by algorithm-recommended content. Holding platforms liable in this way would not require new laws, but merely a recognition of an uncomfortable truth: social media algorithms incentivize

[174] OFF. OF THE N.Y. STATE ATT'Y GEN. LETITIA JAMES, *supra* note 2, at 4. The report also goes on to advocate for reform of CDA § 230 to "require an online platform to take reasonable steps to prevent unlawful violent criminal content (and solicitation and incitement thereof) from appearing on the platform in order for it to reap the benefits of Section 230." *Id.*

[175] Brief for Respondents, at 4, Gonzalez v. Google, LLC, 2 F.4th 871 (2023) (No. 21-1333), https://www.aclu.org/wp-content/uploads/legal-documents/gonzalez_v._google_scotus_amicus_brief.pdf [https://perma.cc/LUW2-XCAG] ("There is no way to visually present information to users of apps or visitors to webpages without making editorial choices that constitute, in plaintiff's terms, implicit 'recommendations.'").

[176] In its amicus brief in *Gonzalez v Google*, the ACLU made a bad faith comparison between YouTube algorithm recommendations and search results on Google. YouTube's algorithm recommendation system is entirely different from the way that Google chooses to order its search results. A true comparison would be if you went looking for an article on "how to attract women," and at the bottom of that article, Google provided you with a series of recommended similar articles based on keywords in your search and on your viewing history, getting increasingly extreme in their content, in an attempt to keep you on Google longer. *See* Brief of American Civil Liberties Union of Northern California & Daphne Keller as Amici Curiae in Support of Respondents, at 9, *Gonzalez* (2023) (No. 21-1333), https://www.aclu.org/wp-content/uploads/legal-documents/gonzalez_v._google_scotus_amicus_brief.pdf [https://perma.cc/LUW2-XCAG].

[177] Brief for Respondents, at 4, *Gonzalez*, 2 F.4th 871 (No. 21-1333) ("If the recommendation implicit in selecting particular material to display is sufficient to negate Section 230 immunity, there would be nothing left of the statute's protection.").

[178] 47 U.S.C. § 230(b)(4).

the publication of extremist content and drive users towards radicalizing content. Perhaps holding platforms liable in this narrow way will slow—or shut off entirely—one avenue of radicalization.