JOURNAL FOR DERADICALIZATION

Spring 2021
Nr. 26
ISSN: 2363-9849

Social Networks and the Challenge of Hate Disguised as Fear and Politics

Adam Klein[a1]
[a]Associate Professor for Communication Studies, Pace University

Abstract
This case study examines social networks as the modern intersections of radical discourse and political extremism. But, as this research will show, extremist content in social networks, even that which has telegraphed violent hate crimes, is seldom communicated in textbook forms bigotry or provocations of violence. Today, the true challenge for social networks like Facebook and Twitter is addressing hate speech that reads more like fear mongering and identity politics, and thus, does not get flagged by monitors. From accounts dedicated to inciting fear over the "threat of immigrants" or "black crime," to groups that form around hashtags declaring that a "#whitegenocide" is underway. These narratives represent the more ubiquitous versions of hate culture that permeate these popular spaces and radicalize cultural discourses happening there. This case study explores how such rhetoric has the same capacity to deliver messages of hate, and even incite violence, by investigating six hate crimes from 2019 that were preceded by social media diatribes. The comparative analysis will show how these examples mostly featured nonviolent expressions of cultural paranoia, rather than avowals of violence or traditional hate speech, thus making them harder to detect by programs seeking out such threats in plain sight. The research then examines the user policies of leading social networks to assess whether their guidelines on hateful and violent content are purposed to address the kinds of language that were espoused by these violent extremists. The study considers the strategies being employed by social networks to expose hateful content of all forms, and the need for more prominent counter narratives.

Article History
Received Oct 16, 2020
Accepted Jan 11, 2021
Published Mar 26, 2021

Keywords: Social Media, Fear Speech, Identity Politics, Shooter Manifestos, Hate Crimes

Social Networks and the Challenge of Hate Disguised as Fear and Politics

In 2018, Facebook CEO Mark Zuckerberg sat before a U.S. Senate committee fielding questions regarding a host of contentious issues for the world's largest social network. At the

[1] Corresponding Author Contact: Adam Klein, Email: aklein@pace.edu; Pace University, Department of Communication Studies, NYC Campus, One Pace Plaza, New York, NY 10038; Twitter: @adgklein

midpoint, Senator Ben Sasse would ask arguably the most important question of the hearing: "Can you define hate speech?" (McArdle, 2018). Amid the amalgam of inquiries, this question from the Nevada senator had just cut to the core matter that social networks find themselves facing: How to define hate inside their communities, and then how to address it. Hate crimes in the 21st century follow an all too familiar blueprint. As a deadly shooting at a Pittsburgh synagogue reminded everyone later that year, it begins with intolerance that escalates into tirade online, before graduating into action in the community. Just prior to opening fire inside the Tree of Life Synagogue, the gunman posted a final diatribe on the social network Gab. "I can't sit by and watch my people get slaughtered. Screw your optics, I'm going in," he declared after venting about the Honduran migrants that were traveling toward the U.S. border (Roose, 2018).

It is true that a pattern of violent extremists telegraphing their intentions online has been a disturbing feature of some recent hate crimes. Elliot Rodger brought his misogynistic venom to YouTube where he vowed "retribution" against women before carrying out a lethal shooting spree through UC Santa Barbara (Pickert, 2014). And Omar Mateen denounced "the filthy ways of the west" and promised "vengeance" for ISIS on Facebook the day he killed 49 people at a LGBTQ nightclub in Orlando (Alexander, 2016). At the 2018 hearing, both the Facebook CEO and Senator easily agreed that any speech that promoted or condoned violence against a group of people should be recognized as hate speech and immediately removed from the network.

But the more complex question of how to classify and address nonviolent hate speech was left unresolved. And that may be the more crucial question to answer today. That is because most hateful rhetoric in social networks does not typify the textbook definition of fighting words directed at a particular community. And even outright expressions of bigotry – racial slurs or cultural ridicule – is rare to find, as users know well which words will get them kicked off their favorite social networks. Research on extremism in social media and political blogs has increasingly found that, rather than overt bigotry, today's online hate speech comes in the shape of fear incitements and identity politics (Byman, 2019; Pohjonen, 2019).

The following research examines how extremists in social media are trading messages of hate and cultural superiority for provocations of fear and cultural paranoia. These inverted expressions of hate have become cleverly encased inside faux-political and fear-driving claims that tacitly deliver the same message: That white western civilization is now a race under siege (Phillips & Yi, 2018). In the monitored spaces of social networks, these messages come in a variety of forms, from conspiracy theories to political dog whistles (Zannettou et al., 2018). This case study will examine how such rhetoric has the same capacity to deliver messages of hate and incitement through provocative pretexts that justify actions of cultural "self-defense." Specifically, the research explores how some recent hate crimes that were preceded by social media diatribes have featured these nonviolent expressions of cultural paranoia rather than avowals of violence, thus making them harder to detect by programs seeking out such threats in plain sight.

The study further explores the role of social networks to consider the policies and actions they are undertaking, and how the promotion of "counter narratives" (Meleagrou-Hitchens & Kaderbhai, 2017) might work to expose these deliberately ambiguous forms of intolerance. Because some users are learning to conceal bigotry in permissible forms like fear and identity politics, shining a spotlight on these surreptitious discourses might be the best prescription for countering them.

In fact, social networks recently adopted similar measures to expose disinformation in the 2020 U.S. Presidential Election. As conspiracy theory groups like QAnon spread fictitious claims that the election had been systematically stolen from President Donald Trump – a claim that would be propagated in social media by the president himself – Twitter and Facebook began to flag these posts as false. The violent insurrection at the U.S. Capitol on January 6, 2021, was premised upon the very same claims, further calling into question the appropriate level of response that social media should adopt when confronting such inflammatory conspiracy theories. Facebook and Twitter ultimately chose to deplatform the social media accounts of President Trump and some 70,000 QAnon-supporters, a decision that would have immediate impact on reducing the spread of falsehoods surrounding the 2020

election (Dwoskin, 2021). However, the networks' actions would also reinvigorate the debate about whether deplatforming is a long-term solution to the greater issue of political extremism in social media (Hutchinson, 2021). While deplatforming may be necessary in certain circumstances, this research argues for social networks to amplify their collective voices as gatekeepers, rather than simply as hosts. Through injecting context and forewarnings about those discourses that thematically sow distrust and disdain for other groups, social networks may help their online communities learn to recognize, reject, and report this form of content.

Online Radicalization: From Stormfront to Social Media

Over the past two decades, research in the processes and forms of radicalization have centered on the Internet as a decentralized space where extremist have come together to share and spread their ideologies. Numerous works in online extremism have studied Stormfront.org, the earliest and most visited community for white nationalist/supremacist exchange. Caren et al. (2012) examined the myriad ways that Stormfront's design strategically fostered a sense of community for members to freely post about the white identity in more mainstream contexts such as politics, popular culture, and even dating. Other studies have observed that while a shared white identity occupies much of the forum discussions, socio-political events often redirect that focus back to the denigration of perceived enemy out-groups (Bliuc et al., 2019). And more recent research has explored Stormfront as a measure for understanding how extremist conversation evolves over time, and thereby potentially accelerates a path toward radicalization (Scrivens et al., 2020).

What many of these studies agree upon is that extremist websites like Stormfront serve a critical function in the process of radicalization by providing a digital echo chamber wherein followers are exposed to only those ideas that validate their preexisting racist belief systems, and thus, the community becomes "a reinforcing agent or accelerant" in the process of radicalization (Von Behr et al., 2013, p. 17). Warner (2010) explains further: "If individuals are only in contact with people they already agree with, there is a danger that their opinions

will polarize and become increasingly radical" (p. 431). Coupled with intolerant themes like those featured on Stormfront, or other radical websites like American Renaissance or the Occidental Observer, studies have found a "collective identity" forms in which members share grievances and even linguistic patterns about perceived out-groups (Bliuc et al., 2019).

Beyond the echo chamber effect, research has also pointed to a gradual transitioning in the collective identity of white supremacists online. Hartzell's (2020) investigation of Stormfront found that its forums were appealing to more mainstream audiences by moving away from white supremacist discourse and toward articulating a "communal belonging, common sense, and pride" in white nationalism (p. 129). The transition from supremacist to nationalist might seem like a minor sematic adjustment but, as this study will explore, it represents a broader, more deliberate campaign to rebrand bigotry in "legitimized" terms. Hartzell writes, "Stormfront members work to construct rhetorical distance between white supremacy and white nationalism by affirming the irrationality of white supremacy while imagining white nationalism as reasonable."

Just as the shifting expressions of identity have suggested the intention of radical groups to enter into a more accepted sphere of public discourse, so has the gradual migration of white nationalist activity from fringe websites like Stormfront, to social networks like Twitter and Facebook. According to watchdog organizations like the Simon Wiesenthal Center (2019), the rise in extremist or hateful sentiment in popular social media has surged in recent years, and been linked to more offline organizational activity. Ganesh's (2020) study of the alt-right, as a growing radical network, underscored how social networks like Twitter are the new spaces where white nationalist discourse is spread and amplified through permissible contexts, focusing on populist grievances that highlight themes of white victimization.

The window of permissible discourse is certainly narrower in social media, as compared to a website like Stormfront, and yet within these mainstream spaces, radical voices are learning to communicate intolerance more covertly. Feffer (2019) describes how the radical right has adopted certain narratives in social media involving issues like immigration, that unite "virulent racists and commonplace conservatives," which has had the effect of

"injecting fringe ideas into mainstream culture" (para. 3). The Anti-Defamation League has also described how anti-immigrant groups have "successfully moved the Overton Window – or the boundaries of what's considered acceptable within political action and public discourse" (para. 2). They attribute the success of these groups' efforts to "mainstream hate" to their use of social media, inside which they "demonize immigrants and get their messages across to the public" (para. 6). Research has also found that users with extremist ideologies have increasingly learned to evade the monitoring eyes of social networks. Ayad's (2019) study on the continued presence of terrorist-affiliated videos on YouTube highlighted some of the tactics their authors use to avoid detection. These include labeling their content as educational, and including links to mainstream websites. Effectively, white nationalists and other extremist groups have steadily been moving beyond the closed echo chambers of websites like Stormfront, and into the open and interconnected spaces of social media, where they are learning to game the system.

Censorship versus Exposure

Today, a search in social media for terms like immigrants, Black Live Matter, Muslims, or Jews will often turn up the same spring-loaded narratives: The alleged threat of Hispanic invaders pouring across the border, or the prospect of black-on-white crime, the infiltration of Sharia law inside the U.S., or the Jewish globalist conspiracy supposedly behind it all. On Facebook, a video that has been viewed over 10,000 times forewarns of a "white genocide" now underway. It features black figures populating the planet and culminates in the image of a white child morphing into a brown child.[2] On Twitter, another popular meme depicts the "Islamic Terrorist Network" spread across a map of the United States as it pinpoints the

[2] The "Fight White Genocide" video, which has been viewed over 10,000 times on Facebook, was retrieved from https://www.facebook.com/fightwhitegenocide/videos/586006391735312/

locations of various Muslim centers.[3] Without calling for violence, these narratives play well inside today's social networks where bigots, veiled behind anonymous user accounts, can openly communicate their intolerance as long as they stay within the lines.

And that brings us back to the challenge and choice presently before platforms like Facebook, Twitter, YouTube, and Instagram. As these companies attempt to refine their algorithms and revise their terms of service to better implicate the types of violence-inciting or bigoted expression that is unwelcome in their spaces, they find themselves struggling with the complexities of modern hate speech. Speaking on the use of artificial intelligence (AI) to root out such content, Zuckerberg conceded:

> Some problems lend themselves more easily to AI solutions than others. Hate speech is one of the hardest, because determining if something is hate speech is very linguistically nuanced. You need to understand what is a slur and whether something is hateful, not just in English, but a majority of people on Facebook use it in languages that are different across the world. (Quinn, 2020)

This dilemma of how to identify hateful and extremist sentiment online has been the focus of much debate. While some studies have highlighted the failings of social networks to carry out comprehensive analyses for detecting extremism (Ayad, 2019), most point to a foundational challenge, which is how to define and recognize hate speech in the first place. Meleagrou-Hitchens and Kaderbhai (2017) write, "Internet companies, governments, and researchers alike are still faced with problems associated with negative measures. What constitutes extreme material? What should be censored?" (p. 58). The authors explored the two most common approaches to confronting online extremism, censorship and exposure. Meleagrou-Hitchens and Kaderbhai explain that with "soft" methods of exposure, in which groups also create counter-narratives that educate others about extremist content, "there

[3] The trending meme of the "Islamic Terrorist Network" in the United States was retrieved from https://twitter.com/LyndaAtchison/status/855268368475013124.

remains both a lack of understanding of how this would occur, or how such narratives could be effectively disseminated" (p. 7). However, the recent surge of harmful disinformation regarding alleged voter fraud in the 2020 election, and the collective response to it, has potentially laid the groundwork for how social networks are learning to counter-message dangerous forms of content at both a micro and macro-level. From the labelling of potentially harmful posts, that provides context for their deceptions, to the high-profile exposure and admonishment of such material in news media and advertising campaigns, social networks have begun to demonstrate a more collective and effective response mechanism.

Still, the other common course of action by social networks continues to be the sweeping removal of toxic material, primarily dangerous forms of disinformation and overt hate speech. In some cases, these deplatforming efforts have proven effective. Alexander (2017) demonstrated how Twitter's suspensions of ISIS-supporting accounts significantly diminished that terror network's "ability to gain traction on the platform, likely hindering their reach to potential recruits" (p. 19). More recently, research conducted by Zignal Labs on Twitter's deplatforming of Donald Trump and QAnon accounts showed a rapid 73% drop in subsequent disinformation regarding "election fraud" (Dwoskin, 2021).

However, several studies have come to similar conclusions on the limitations and long-term ineffectiveness of censorship: that simply closing the door to one social network will lead radical users to the next venue where fewer restrictions exist. Addressing this resiliency factor, Johnson et al. (2019) write:

> [T]he key to understanding the resilience of online hate lies in its global network-of-network dynamics. Interconnected hate clusters form global 'hate highways' … Our mathematical model predicts that policing within a single platform (such as Facebook) can make matters worse, and will eventually generate global 'dark pools' in which online hate will flourish. p. 261.

Many such 'hate highways' have arguably already been formed, leading to those communities that advertise fewer constraints on speech. Zannettou et al. (2018) and Comerford (2020) each observed a high prevalence of hate speech forging inside less regulated domains like 4chan, 8chan, and Gab, which advertise themselves as a welcoming alternative to the mainstream and restrictive social networks.

The present research does not argue that either one of these two approaches, censorship or exposure, is always the correct remedy for countering extremism. There are different contexts that should be considered, such as whether the material in question promotes violence. However, where bigotry takes on permissible forms like fear-mongering or identity politics in order to obscure the radical ideologies behind them, a counter-narrative approach might better educate the public about how such rhetoric is associated with hate speech, thus priming them to identify and reject it. As this study will show, some assailants have used these gray areas of extremism in social media to espouse their "political" and "fear-based" justifications for committing hate crimes, underscoring the need to further establish these style communications as vehicles for hate.

The Potency of Fear

When David Duke ran for the US Senate in Louisiana in 1990, the former Ku Klux Klansman never used racial slurs when speaking of African-Americans, who were nonetheless a centerpiece of his campaign. In fact, according to those that covered his Senate run, Duke "rarely even used the word black," because he had mastered the use of "other codes" like his references to "the massive rising welfare class" and "welfare mothers who just have babies," or the perils of affirmative action (Maraniss, 1990, para. 9). More than being adept at dog whistle politics – the practice of speaking in coded terms that those attuned to racism would understand – Duke had obtained a shrewd understanding of how to employ "racial threat" as a mobilizing idea (Giles & Buckner, 1993). This was 1990, and the days of segregation platforms or assertions of racial superiority were relics that would not serve him in this

political era. But racist sentiment was still a potent messenger if one knew how to resurrect it. Duke skillfully appealed to the fears of a sector of the white working-class that felt threatened by the rising black population (Boulard, 1991). He also knew how to lay blame at the doorstep of another scapegoat, the Jewish community, again without engaging in direct anti-Semitism. Maraniss (1990) writes:

> At a rally in Shreveport, Duke went after one of his favorite targets, the New Orleans newspaper. "You know the New Orleans Times-Picayune is down on me," Duke told supporters. "But you know who owns the Times-Picayune, don't you?" From the front row came the reply: "Jews. Jews. Jews." (para. 15)

Duke's capacity to harness hate inside the electorate was not just about his crafty employment of euphemisms and insinuation. Rather, it was about the power of using fear rhetoric and identity politics to cultivate the same level of bigotry from receptive audiences. Why do these rhetorical forms work, and how do they flourish inside today's social networks?

Let us consider the hashtag #whitegenocide, which a 2016 study cited as the most tweeted phrase among white nationalists (Berger, 2016). It has since come to encapsulate an array of messages pertaining to an "endangered white race," as further research has showed by tracing the hashtag through Twitter dialogues on Hispanic immigration, Muslim refugees, and interracial crime (Deem, 2019). There are three reasons why #whitegenocide has become a prevalent device among white nationalists and other authors of bigotry. First, it allows these groups to integrate their extremist message into the body politic of social networks. By marrying racial resentments to mainstream politics, as Duke did with affirmative action in the nineties, racists are attempting to *legitimize hate* inside the political exchange of social media. And the coalescence of #whitegenocide with timely topics, like border security, provides a certain degree of cover to those wishing to espouse bigotry in a context that is accepted in social media (see Figure 1). Second, these stories of racial threat allow racists to adopt the more accessible stance of "victim," which in turn suggests a position of *justifiable outrage*. It

is here where a troubling pattern has emerged among those who have transferred the message of white victimization into a rationale for violence (see Figure 2).

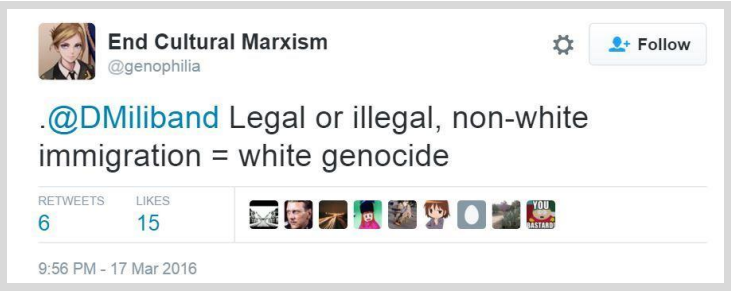

Figure 1. Linking "white genocide" to the issue of immigration.

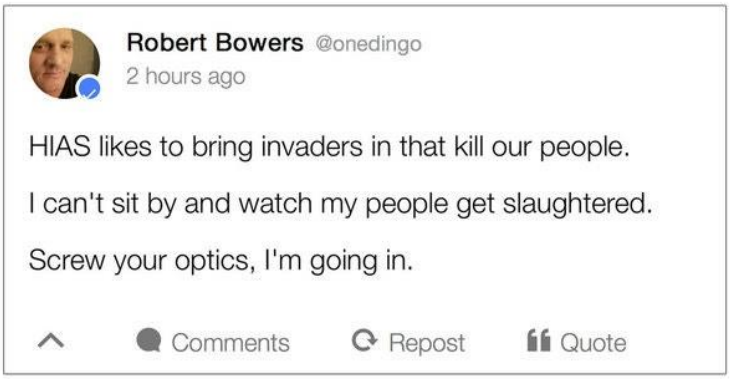

Figure 2. Robert Bowers' final post prior to his deadly attack on a Pittsburgh synagogue.

The third motivation for trading the old declarations of 'white supremacy' for today's outcries of 'white genocide' is that, where the former once fostered a false sense of superiority, the latter *produces fear*. And research has shown that fear can be a potent deliveryman for hate. Gerbner's (1998) theory of the Mean World Syndrome was a critical branch of his larger work in Cultivation analysis; studying how the heavy viewing of television led audiences to perceive their worlds as a reflection of the TV world. In terms of the superfluity of violence in television news and entertainment, the Mean World Syndrome

suggested that the steady consumption of such content cultivates in audiences a sense of fear about the likelihood of violence happening to them, as well as a deep paranoia about those groups traditionally depicted as the culprits:

> Gerbner's bottom-line point was that without positive representations to balance the bad, the meanest members of minority groups are allowed to stand in for all the rest – creating a distorted and menacing picture that leaves viewers feeling under attack, and reinforcing a siege mentality that feeds and feeds off of anger and rage. (Morgan, 2010)

Of course, today, narratives of imminent danger can be found in all forms of new media, from online news, to political blogs, to the hostile representations of our world that we share with each other in social networks. And so returning to the messages of cultural threat that populate these networks, especially where political talk is happening, the perception of a white genocide underway is just one of a series of fear-driving claims that circulates within this marketplace of ideas. Others that will be examined in this case study include the professed threats of Hispanic migration, Muslim refugees, and Jewish control.

Gerber's point was that fears brought on by exposure to such messages of cultural threat and notions of violence happening to us, eventually lead to another emotion – hostility. It is for this reason that statements about the threat of refugees or immigrants have become the preferred vehicle for racists in communities across social media. There, they can parade as politics, while in fact stoking cultural anxieties that can lead to blame and bigotry.

Violence Between the Lines

In their study of online discourses surrounding the migration of Syrian refugees into Europe, Sayimer and Derman (2017) explored the prevalence of "fear speech" as a deeper, and potentially more dangerous, form of hate speech. Analyzing anti-refugee videos on YouTube,

they describe how fear "rationalizes and legitimizes racism" in this shared context, and in doing so, can provide certain viewers with the validation they are seeking to take action. They wrote, "Violence coined as 'self-defense' seems more acceptable" (p. 392). Here, fear-based hate speech can be understood as either a genuine form of paranoia expressed by hysterical individuals, or a useful ruse whereby bigotry is being dressed up as fear in order to justify one's hate. In both cases, fear is connected not only to hate, but also to the potential for violence. Buyse (2014) explains, "The instigation of fear among one's own group, rather than hatred against the other, has been found to be a key mechanism in such processes leading to violence. This may lead to accepting state violence against the group one fears" (p. 785).

For social networks, whose chief concern is stemming that content which could inspire or even preview hate crimes, the challenge returns to language. When does an authentic criticism over a subject like the influx of foreign refugees cross a threshold into becoming a pretext for justifying hostility and violence? For companies like Twitter, the complex task of determining the intent and character of a hateful commentary includes the question of whether the words should be taken seriously, or even attributed to the author. After all, much of what we find in extremism in social media comes couched in the form of humor, or opinions that seem twice removed from their author, such as a retweet that shares the bigoted ideas of others. There are reposted news headlines that, strung together, are meant to highlight the offense of one particular group, or memes that demagogue entire populations. And then there are the more extreme examples of cultural rants that have signaled the violent intent of their authors. These cases are seldom on the radar of social networks until after the assailants commit their acts. Thus, social networks continue to contend with the ambiguity of radicalism as it arises in their spaces, taking on its many forms, from fear, to humor, to political commentary. Many political leaders have called on social networks to swiftly curtail this toxic trend, as studies continue to show a correlation between online radicalism and offline violence (Comerford, 2020; Wojcieszak, 2009). One landmark study recently found a strong connection between increased Facebook use in towns in Germany and anti-refugee hate crimes occurring in those same areas (Müller & Schwarz, 2018). But to see the problem up

close, and from the perspective of those charged with resolving it, one must first analyze some of the actual messages that precede hateful violence.

Case Study

The following case study investigates six social media publications whose authors went on to commit deadly hate crimes in the hours or days that followed their final posts. Our attempt is to understand, as social networks must do, the patterns that potentially exist among these communications. A comparative analysis was performed on the social media content of the assailants, focusing on the rhetorical nature of their communications and whether they expressed messages of hate speech, violence, fear, political commentary, or other forms of declaration. The examples selected were premised on hate crimes that occurred in 2019, carried out by assailants whose collective violence stole 93 innocent lives that year. The research analyzed their online writings prior to these acts. These communications varied in length and form, including multipage manifestos, public Tweets and Facebook posts, as well as obscure comments left on websites like YouTube and Instagram. The study then returns to the social networks' user policies on hate speech and violent content to assess whether such guidelines are purposed to address the kinds of language that these violent extremists expressed.

The Most Extreme Cases and the Shape They Take

For a moment, consider the geography of a social network from the perspective of someone wishing to espouse extremism there. Why do social networks present fertile ground for these actors? Often, anonymity is cited as a central factor in allowing intolerant individuals to air the kinds of viewpoints they would otherwise never share in public (Brown, 2018). Indeed, whereas the 2017 white nationalist gathering in Charlottesville was televised and highly revealing, social networks on the hand can be discreet, and therefore highly disinhibiting for the undeclared racist. But beyond concealment, these sites provide something

else: listeners. Whether real or imagined, social networks present the sense that an extremist's thoughts are finally being heard, and even occasionally validated in the form of likes and retweets. Among the following cases of persons who went on social media before delivering violence to the public, they share notably different ideologies and targets. And we will see how difficult it is to establish a clear code of communication among them, the kind that could potentially thwart future threats. But one quality that intersects these assailants' final posts was their motivation for a captive audience, and that element will be critical later when considering measures to combat hate culture in a medium that continues to present this stage.

"He's followed the formula"

For Brenton Tarrant, the stage of social media was literally that place from which he broadcast his rampage. Tarrant's mass shooting spree in March of 2019 was directed at two mosques in Christchurch, New Zealand, and was also livestreamed to a global audience on Facebook. He claimed 51 lives. Like the other cases that will be examined here, Tarrant's action was a hate crime preceded by social media activity that either directly or indirectly previewed these ensuing events. As mentioned, all of these attacks were staged inside a single year. But Tarrant's crime stands out for two reasons. First, it predated the others and was thus emulated in form by some of those shooters that followed. Speaking on Tarrant's celebrated status among white nationalists, terrorism expert Greg Barton said, "He's followed the formula and done what the formula delivered" (Maley, 2019, p. 3). Secondary to Tarrant's influence was the role that social media played in the Christchurch massacre as both that of his publisher and broadcaster. On the morning of March 15, the 28-year-old Tarrant posted a link on Twitter that brought his followers to a 74-page manifesto on 8chan, where he spelled out his motivations and alleged justifications for the oncoming violence. Minutes later he carried out his massacre, which he simultaneously livestreamed on Facebook for all to see.

Analysis of Tarrant's manifesto, "The Great Replacement," titled after a French novel bearing similar themes, evidences a mind that was fixated on the notion that the white identity is being erased by a process of mass migration into Western countries. Feffer (2019) explains,

"The idea of "the great replacement" is based on the fantasy that "they" (especially migrants and Muslims) are intent on replacing "us" (whites, Christians)" (para. 5). For Tarrant, his writings present a dire manifestation of the white genocide mantra examined earlier, elevating it to justification to kill as many Muslims as possible and inspire others to do the same. Much of Tarrant's manifesto is presented in Q&A style, in which he poses the questions he assumes others will ask, and then answers them. "Were/are you are you a neo-nazi?"[4] No, he responds under the assumption he might be killed during the attack. In fact, he was not. "Do you feel any remorse for the attack?" No, he writes, "I only wish I could have killed more invaders." "Why did you target those people?" Tarrant addresses this question often throughout his manifesto, though seldom enters into direct insults of Muslims or racial slurs on Islamic culture. He even contends he does not hate Muslims who live in their homelands, but dislikes those "choosing to invade our lands live on our soil and replace our people." Tarrant's obsession with his white race becoming a "demographic minority" is presented as his central motivation:

> It's the birthrates. It's the birthrates. It's the birthrates. If there is one thing I want you to remember from these writings, it's that the birthrates must change … Even if we were to deport all Non-Europeans from our lands tomorrow, the European people would still be spiraling into decay and eventual death.

It is from this paranoia that Tarrant pivots to self-aggrandizing expressions of martyrdom. He casts himself as nobly justified to "take a stand to ensure a future for my people."

The citing of pretexts for taking defensive action, such as "taking a stand for our people," is significant because such positions can be found throughout social media today in very much the same context as Tarrant is using here. Of course, in this lethal framework, we

[4] Brenton Tarrant's manifesto was reviewed through the Centre for Analysis of the Radical Right and at https://crimeresearch.org/wpcontent/uploads/2019/03/The-Great-Replacement-New-Zealand-Shooter.pdf.

look back at the writing that preceded a massacre and believe that what we are seeing is the very formula that social networks are trying to identify: A racist ideology, expressed in clearly searchable terms like "Muslim invaders," "Non-Europeans," and "take a stand," that foreshadows violent intent. And, to the extent to which Tarrant expressed these delusions, in 74-pages, and alluding to his desire to attack, there is the impression that such waving red flags can be detected in social media to avert future violence. But that impression would be an illusion. Realistically, little could be done to prevent these attacks, for only minutes after Tarrant posted the obscure Twitter link, he engaged in his assault at the Al Noor Mosque and Linwood Islamic Centre.

Moreover, Tarrant's words were not exceptional by today's standards of online white nationalist rhetoric, especially if one visits conspiracy-laden websites like 8kun, Reddit, Gab, and even Twitter, where political sentiments about standing up to so-called "invaders" are common. But the Christchurch massacre carried one additional element of social media activity that now presents a new problem for social networks. As described, Tarrant filmed his shooting spree and streamed it live over Facebook, where the deadly images were not only seen but also captured by viewers, only to later be replayed again and again:

> Copies of that footage quickly proliferated to other platforms, like YouTube, Twitter, Instagram and Reddit, and back to Facebook itself. Even as the platforms worked to take some copies down, other versions were re-uploaded elsewhere ... "It becomes essentially like a game of whack-a-mole" says Tony Lemieux, professor of global studies and communication at Georgia State University. (Perrigo, 2019, para 2)

Tarrant's use of social media to disseminate both his ideology and his horrific acts offers a critical insight into how such audience-intended hate content can not only spread, but also radicalize other potential extremists, even activating in some a desire to follow.

"Inspired by the horrific acts"

While Tarrant's scenes of bloodshed began to emerge in the open plain of social media, they offered a vivid glorification of hateful violence that was evidently alluring to likeminded individuals. In the months that followed, two new shooters surfaced who cited the Christchurch carnage as their inspiration, and in doing so, followed a recognized pattern of extremists who "copy behaviors observed directly or in the media" (Youngblood, 2020, p. 2). John Earnest, 19, attempted a killing spree at a synagogue in Poway, California, after posting his own manifesto to 8chan, writing, "Tarrant was a catalyst for me personally. He showed me that it could be done. And that it needed to be done."[5] Earnest also attempted to livestream his hate crime on Facebook, but was not successful. His rifle jammed amid the attack and he only managed to fire off a few rounds, injuring worshippers, but tragically killing one woman, Lori Gilbert, who threw herself over the Rabbi to shield him.

Three months later, 21-year-old Patrick Crusius also followed in the footsteps of Brenton Tarrant, again posting a manifesto to 8chan minutes before entering a Walmart in El Paso, Texas with a semi-automatic. His online declaration began, "I support the Christchurch shooter and his manifesto. This attack is a response to the Hispanic invasion of Texas. They are the instigator, not me."[6] Where Tarrant and Earnest directed their rage at the Muslim and Jewish communities, respectively, Crusius targeted Hispanic Americans and immigrants, killing 23 people. The Crusius manifesto read like a copy of Tarrant's declaration, sounding alarms about the "Great Replacement" of Americans, with Crusius now assuming the role of heroic martyr: "I am simply defending my country from cultural and ethnic replacement brought on by an invasion." His hate was also embedded in identity politics, as evidenced not

[5] John Earnest's manifesto was reviewed through Middle Eastern Media Research Institute's Domestic Terrorism Threat Monitor, and at https://archive.org/details/JohnEarnestManifesto_201905/page/n3/mode/2up.

[6] Patrick Crusius' manifesto was reviewed through the Counter Extremism Project and at https://grabancijas.com/patrick-crusius-manifesto-the-inconvenient-truth/.

only in his manifesto, but his digital footprints that included several Twitter posts celebrating President Trump's planned border wall and tough stance on Hispanic migrants.

The parallels between these shooters who were inspired by Tarrant's writings extended to his favored themes of fear, blame, and cultural preservation. Earnest wrote about the dangers of the "tyrannical and genocidal Jew," warning readers, "You should be more afraid of losing your entire race than this life you now live." Crusius echoed the fixation on "invaders" having "the highest birthrate," cautioning that immigration would lead to America's undoing. Both offered themselves up as cultural martyrs of sorts, from Crusius who called on his readers to keep up the fight should he die, declaring, "I am honored to head the fight to reclaim my country from destruction," to Earnest who said, "I only wish to inspire others and be a soldier that has the honor and privilege of defending his race." One journalist later summarized how Earnest was "inspired by the horrific acts that preceded him: mass shootings at a synagogue in Pittsburgh and at two mosques in New Zealand" (Parvini, 2019, para. 1). Disturbingly, both he and Crusius indicated a desire to now do the same, and in all likelihood succeeded in transmitting their words to a new set of listeners.

"There's not really one profile"

Where the carefully constructed statements of the first three shooters took on a very similar shape and agenda, albeit different targets, the latter three cases were notably dissimilar in message and motivation. Connor Betts, a 24-year-old from Dayton, Ohio, shot and killed nine people at a local bar, including his sister. Later, those that knew Betts said he had a long history of misogyny and once reportedly compiled a "rape list" in high school (Grady, 2019, para. 2). His August 4th shooting followed months of social media posts in which he professed his ardent support for socialism, presidential candidate Elizabeth Warren, and guns. He also railed against everything from Republicans to moderate Democrats, to oil executives, white nationalists, and even the recent mass shooter, Patrick Crusius. But while Betts cultivated a politically charged mindset, his writings were not exactly the blatant expressions of political extremism. He used Twitter to share declarations like, "I want

socialism, and i'll not wait for the idiots to finally come round to understanding" (Croucher, 2019, para. 10). And, "Millenials have a message for the Joe Biden generation: hurry up and die," which he posted just hours before the shooting. In fact, if there was anything remarkable about Betts' social media tirades, it is how unremarkable they were. Beyond some darker posts, like "kill every fascist," which he tweeted about a white nationalist rally, or retweeting one post that satirically called for the beheading of oil executives to combat climate change, the Twitter account of Connor Betts resembled that of many other outrageous political commenters.

Mohammed Alshamrani was more direct in his use of social media to explain his contempt for America as "a nation of evil," though he did not forecast his violent intentions as others did, and his Twitter activity could hardly be called a manifesto. The 21-year-old Alshamrani was a visiting aviation student from Saudi Arabia, training at the Naval Air Station in Pensacola, Florida. On December 6, he opened fire in one of his classrooms, killing three Navy sailors and wounding others. His violence was presumed to be an act of terrorism, and tellingly, some of the other Saudi students filmed Alshamrani's massacre. They were detained after he was killed. Beyond his terrorist designation, Alshamrani's operation should also be understood as an act of hate. His abhorrence was directed at Americans, and the sailors he killed were also his classmates. But on Twitter, he sought to present himself as neither culturally nor ideologically motivated: "I'm not against you for just being American, I don't hate you because your freedoms, I hate you because every day your supporting, funding and committing crimes not only against Muslims but also humanity. I'm against evil" (Allen, 2019, para. 3). He went on to express disdain for American support of Israel, but beyond his status as a visiting member of the Saudi Air Force, his tweets were probably no more conspicuous than others critical of U.S. policy in the Middle East or Israel. And though ominous, his words were not openly belligerent. "You will not be safe until we live it as reality in [Palestine], and American troops get out of our land," he posted prior to his violence.

Lastly, only four days later, David Anderson and Francine Graham carried out another hate crime, this time in New Jersey. The middle-aged couple, followers of the Black Hebrew Israelite fringe movement, began their shooting spree at a Jersey City cemetery. They killed one law enforcement officer and then proceeded to a nearby kosher supermarket, where inside, they executed three civilians before being killed themselves in a standoff with police. Anderson, viewed as the mastermind behind the attacks, had been active on several platforms including Instagram, YouTube, and SoundCloud. Using the pseudonym Nawada Maqabath, he seethed for months about the Jewish "imposters" and "wicked Israelites" who were supposedly inducing police violence against the Black community: "These police are NOT scared, they are carrying out a well planned agenda layed out by the upper echelon of Rosenbergs people" (McBride, 2019, para. 21). On YouTube, Anderson also expressed an affinity for Gavin Long, who had himself killed police officers in 2016 after sharing extremist conspiracies on YouTube. Echoing Long, Anderson wrote, "There are NO innocent cops. The entire organization was built with nefarious intentions against Israelites." Because his writings were published under a pseudonym and spread across multiple sites and comment sections, Anderson's radicalism was like most online in that it was obscure, sporadic, and thus, went undetected in the ether of the Internet.

As we next return to the question of how social networks learn to locate these sources of hate inside their communities, it is important to recognize where the thread line exists between these incidents, and where it does not. Tarrant, Earnest, and Crusius' manifestos bore a notable resemblance in style and tone, but the examples of Betts, Alshamrani, and Anderson illustrate the inconsistent nature of these communications. Betts, who was potentially motivated by misogyny expressed political zealotry in his tweets – "kill every fascist" – but in truth, no more severe than the language one finds in countless social media texts. Alshamrani was inspired by his hate for Americans, though on Twitter his posts were temperate, assuming the more reasoned stance of, "I'm against evil." And Anderson used religious grounds to justify his anti-Semitism, characteristic of the kind of conspiracy rhetoric that makes its author out to be the victim of his world, rather than its next assailant.

All of these varied examples come back to the underlying question: Should social networks have been able to flag these dangerous actors?

A study of every mass shooting since 1966 found that while certain commonalties do exist among violent extremists, such as a desire for notoriety, the assailants are often more distinct than we realize: "What our research is starting to uncover is there's not really one profile of a mass shooter" (Pane & Dazio, 2019, para. 9). It is therefore more critical to consider the elements of language, rather than the profiles of shooters, in attempting to isolate hateful extremism from within the vast traffic of social networks.

Social Networks as Gatekeepers

In many ways, the challenge before social networks is not unlike the test that many universities now face when a known extremist has been invited to speak on their campus: To host or not to host? Like universities, social networks understand that opening their doors to the likes of a Richard Spencer, someone who has called for "peaceful ethnic cleansing," or Louis Farrakhan, who recently compared Jews to termites, lends the credibility of their community to these provocateurs of racism and anti-Semitism. And beyond credibility, these networks also provide unprecedented followings and approval to the kind of sentiments that would normally be denounced if they were carried on banners. Yet currently, Spencer and Farrakhan each have prominent accounts on Twitter because, in spite of their widely known beliefs, they have not broken the social network's terms of service. So then, what kinds of rhetoric is prohibited in these spaces, and do these rules of communication and conduct come close to addressing the kinds of hate speech that we have been exploring?

An analysis of the user policies of leading networks reveals a common approach to defining hate speech. Facebook, for example, classifies hate speech along three tiers from the most extreme, "violent or dehumanizing speech," to "statements of inferiority," to "calls for exclusion or segregation" (Community Standards, 2020). Each of these categories pertains to language directed at groups with "protected characteristics," such as race, gender, and

ethnicity. Instagram, which is owned by Facebook but operates independently, forewarns users that it is not a site for expressing "support or praise" for hate groups. They, too, classify hate speech as the kind of language that promotes violence or levies attacks on someone based on their identity (Community Guidelines, 2020). YouTube continues this trend, chiefly categorizing hate in terms of expressed violence against groups or content that in any way dehumanizes, alleges superiority or inferiority, or uses racial or religious slurs (YouTube Policies, 2020). Notably, YouTube also forbids the use of conspiracy theories to malign others, including the denial of "well-documented violent events, like the Holocaust." This is significant because it represents one of the few instances of a social network recognizing the use of coded language as a vehicle for hate. But what these policies do not address are the ubiquitous expressions of fear, specifically fear of minorities, which effectively serves to foment feelings of hate for them.

Returning to the six extremists whose lethal acts were preceded by social media tirades, we speculated whether their communications embodied the sort of characteristics that would be readily classified as hate speech by these social network guidelines. And while some of the language in the manifestos spoke of taking action, the need to defend the country, or even praise for recent hate crimes, most of the comments fixated on the perceived threat of others. By and large, these extremists used a rhetoric of fear and the façade of politics to cast blame on those communities that later became the subject of their violence. But despite this overwhelming theme, Facebook, Instagram and YouTube's continued policies on hate speech do not account for the pervasive provocations of fear.

Twitter, by contrast, represented one of the only social networks to address fear as an inverted form of hate. They state, "We prohibit targeting individuals with content intended to incite fear or spread fearful stereotypes about a protected category, including asserting that members … are more likely to take part in dangerous or illegal activities, e.g., "all [religious group] are terrorists" (Twitter Rules, 2020). Yet, notwithstanding Twitter's more encompassing definition of hate, it is unclear how this policy is presently being enforced in

this space where fear-based bigotry is rampant as demagogic hashtags like #WhiteGenocide, #BlackCrimes, and #MigrantInvasion spread like ivy through their pages.

In the face of this proliferation of hate, social networks have turned to a natural ally – technology. Artificial intelligence programs are being employed to seek out the proscribed forms of hate speech, that they may flag, and sometimes, eliminate such content as it surfaces. But even if AI software is given an exhaustive list of violent terminologies and dehumanizing phrases to locate, we have seen how modernized hate speech can be well disguised. It appears political, humorous, and even creative, all toward the goal of traveling below the radar. Moreover, opinion surveys have found that while a majority of social media users want social networks to remove hateful content (Edwards-Levy, 2018), most citizens cannot agree on what actually constitutes hate speech, as opposed to fair political comment (Ekins, 2017). This notable discrepancy begs the question on the issue of censorship: How can one expect an algorithm to discern hate speech if we ourselves cannot?

Thus, the expectation for social networks to fully purge their digital environments of hate is neither practical nor likely. In these densely trafficked spaces, the "game of whack-a-mole" will likely continue in which one toxic thread is taken down, only to be replaced by another. And, if a hateful ideology is effectively blocked from residing in one social network, earlier studies established that these radical narratives simply migrate to another website with fewer restrictions (Comerford, 2020; Johnson at al., 2019; Zannettou et al., 2018). It is therefore perhaps more counteractive for social networks to begin thinking of ways to reform their communities in order to make them less inviting for racists to occupy.

Rethinking Hate in Social Media

For social networks, the adoption of more nuanced understandings of hate speech could allow them to better track the signs of intolerant authorship where it begins. Beyond provocations of violence, racial slurs, and other explicit forms of bigotry, this research investigated implicit forms of hate that permeate political discussions and toxify our forums.

These can include incitements of fear over the threat of minorities ('Muslims are plotting to attack us'); political tirades that encircle issues of identity ('Immigrants are taking away our country'); cultural conspiracy theories ('Jews are funding a migrant invasion'); and language that speaks of heroic resistance against a designated enemy ('Defend your religious freedom against the gay rights agenda'). Such narratives, common in today's social networks, are often smokescreens for bigotry, and have sometimes been used as pretexts for violence.

Until recently, social networks have been reluctant to extend their roles as gatekeepers in ways that would editorialize or rebuke the free speech of their users. But with the extraordinary surge of disinformation surrounding the 2020 election, social networks took unprecedented steps to begin exposing the deceptive content, while also taking to the airwaves to vociferously speak out against its future proliferation inside their communities. This counter narrative response was also applied to misleading information on COVID-19 and voting during the election. Simultaneously, the CEOs of Facebook and Twitter became active in national media appearances to publicly reject any notion that false narratives on public health or voting was welcome in their spaces (CBS News, 2020).

That same combination of a counter-messaging could be used to address the covert forms of extremism that bypass the community guidelines of social networks but nonetheless infiltrate and corrupt social discourses. At the micro-level, labeling, explaining, and exposing such material for what it is doing – inciting fear, spreading conspiracies, or engaging in identity politics – could stigmatize these offenses, as it inoculates and educates the community against them. And at the macro-level, a high profile and prolonged denunciation of identity politics and fear-incitement in social media may produce a counter narrative that reinforces the public's recognition that these *are* forms of extremism. The strategy behind counter-narratives is that they represent the positions of the platforms, but also give voice and purpose for "empowering online actors to engage and denounce extremist propagators" (Meleagrou-Hitchens & Kaderbhai, 2017, p. 53).

Finally, platforms like Twitter, Facebook and YouTube must consider how to address the stage factor in which extremists see these communities as ideal places to finally air their

personal vitriol before a receptive audience. For extremists, the social factor of social media is critical because they have come to these networks to indulge their intolerant views with the belief that others are listening, and agreeing with them, and sometimes they are right. In the most dire cases, some would-be extremists observed the pattern by which assailants of hate crimes were able to gain notoriety, not only from their violent deeds, but from the words they published in social networks which were then immortalized in the media following their assaults. Thus, they too seek out this infamy following that same blueprint.

Social networks must examine new ways to stem the notion that hostile and fear-driving rhetoric about different groups is somehow welcome in their spaces. The repeated public exposure and resounding denunciation of such discourses might be the best prescription to deterring hateful diatribes. Perhaps the clearest evidence of the effectiveness of community opposition to malicious posting can be found in examining those networks that impose *few* restrictions and renunciations of belligerent speech. Platforms like Reddit, Gab, and formerly 8Chan, embraced a more open policy of communication, and thus quickly became the preferred homes to more virulent authors, unfiltered hate speech, and shooter manifestos. What does this tell us? Perhaps that advocating for a community that discourages hateful content matters, because it sets the tone and communication climate that other users will feel empowered to protect.

Future research might explore how social networks are educating their communities about hate content, both in terms of defining it and discouraging its expression. A focus on Twitter's recent identification of fear-inciting bigotry will be instructive to learn how that network is addressing this more concealed form of hate where other networks are not. But while revamping user policies on hate speech is a central step to isolating the language of extremists, such measures serve little purpose if the community does not know about them. Equally, the efforts to reduce hate speech in social media can only achieve a surface-level success, at best, if we just choose to recognize the hate that comes in obvious forms.

References

Alexander, A. (2017). *Digital Decay: Tracing Change Over Time Among English-Language Islamic State Sympathizers on Twitter.* The George Washington University Program on Extremism. Retrieved February 14, 2021, from https://extremism.gwu.edu/sites/g/files/zaxdzs2191/f/DigitalDecayFinal_0.pdf

Alexander, H. (2016, June 16). Omar Mateen searched for Facebook posts about Orlando shooting while he was carrying it out. *The Telegraph.* Retrieved May 1, 2020, from https://www.telegraph.co.uk/news/2016/06/16/omar-mateen-searched-for-facebook-posts-about-orlando-shooting-w/

Allen, N. (2019, December 7). Saudi gunman called US 'nation of evil' before Florida shooting, monitoring group says. *The Telegraph*. Retrieved June 1, 2020, from https://www.telegraph.co.uk/news/2019/12/07/saudi-soldier-kills-four-wounds-eight-us-naval-base-shooting/

Anti-Defamation League (n.d.). Mainstreaming Hate: The Anti-Immigrant Movement in the U.S.. Retrieved May 1, 2020, from https://www.adl.org/the-anti-immigrant-movement-in-the-us

Ayad, M. (2019). *'The Baghdadi Net': How A Network of ISIL-Supporting Accounts Spread Across Twitter*. London: Institute for Strategic Dialogue.

Berger, J.M. (2016). *Nazis vs. ISIS on Twitter: A comparative study of White nationalist and ISIS online social media networks.* George Washington University Program on Extremism. Retrieved May 1, 2020, from https://www.extremism.gwu.edu/sites/g/files/zaxdzs2191/f/downloads/Nazis%20v.%20ISIS.pdf

Bliuc, A.-M., Betts, J., Vergani, M., Iqbal, M., & Dunn, K. (2019). Collective identity changes in far-right online communities: The role of offline intergroup conflict. *New Media & Society, 0*(0). doi:10.1177/1461444819831779

Boulard, G. (1991, June 13). Former KKK Leader Duke Attacks Affirmative Action. *Christian Science Monitor.* Retrieved May 1, 2020, from https://www.csmonitor.com/1991/0613/13072.html

Brown, A. (2018). What is so special about online (as compared to offline) hate speech? *Ethnicities, 18*(3), 297–326. doi:10.1177/1468796817709846.

Buyse, A. (2014). Words of Violence: "Fear Speech," or How Violent Conflict Escalation Relates to the Freedom of Expression. *Human Rights Quarterly, 36*(4), 779–797.

Byman, D. (2019). The Threat From Within. *National Interest, 162*, 52–60.

Caren, N., Jowers, K., & Gaby, S. (2012). A social movement online community: Stormfront and the white nationalist movement. *Research in Social Movements, Conflicts and Change, 33*, 163-193.

CBS News (2020, September 3). "Mark Zuckerberg: Flagging misinformation about mail-in-voting "will apply to the president." Retrieved October 1, 2020, from https://www.cbsnews.com/news/facebook-zuckerberg-2020-election-misinformation-president-trump/

Comerford, M. (2020). Confronting the Challenge of Post-Organisational Extremism. Retrieved October 1, 2020, from https://www.orfonline.org/expert-speak/confronting-the-challenge-of-post-organisational-extremism/

Community Guidelines (n.d.). Instagram. Retrieved February 22, 2021, from https://help.instagram.com/477434105621119

Community Standards (n.d.). Facebook. Retrieved February 22, 2021, from https://www.facebook.com/communitystandards/hate_speech

Croucher, S. (2019, August 6). Twitter Account Linked to Connor Betts Suggests Dayton Shooter was Democrat-Supporting Satanist. *Newsweek.* Retrieved May 1, 2020, https://www.newsweek.com/connor-betts-twitter-dayton-shooting-socialist-satanist-1452719

Deem, A. (2019). The Digital Traces of #whitegenocide and Alt-Right Affective Economies of Transgression. *International Journal of Communication, 13*, 3183–3202.

Dwoskin, E. (2021, January 17). Misinformation dropped dramatically the week after Twitter banned Trump and some allies. *Washington Post.* Retrieved February 20, 2021, from https://www.washingtonpost.com/technology/2021/01/16/misinformation-trump-twitter/

Edwards-Levy, A. (2018). Most People Say Social Media Sites Should Crack Down On Harassment, Fake News: Poll. *Huffington Post.* Retrieved August 1, 2020, from https://www.huffpost.com/entry/social-media-harassment-fake-news-poll-alex-jones_n_5b7b1c53e4b0a5b1febdf30a

Ekins, E. (2017, November 8). *82% say it's hard to ban hate speech because people can't agree what speech is hateful*. CATO Institute. Retrieved May 1, 2020, from https://www.cato.org/blog/82-say-its-hard-ban-hate-speech-because-people-cant-agree-what-speech-hateful

Feffer, J. (2019, October 21). Far Right Extremism Has Gone Mainstream. Retrieved June 1, 2020, from https://ips-dc.org/far-right-extremism-has-gone-mainstream/

Ganesh, B. (2020). Weaponizing white thymos: flows of rage in the online audiences of the alt-right. *Cultural Studies*. doi: 10.1080/09502386.2020.1714687.

Gerbner, G. (1998). Cultivation Analysis: An Overview. *Mass Communication & Society,1*(3), 175–194. doi:10.1080/15205436.1998.9677855.

Giles, M., & Buckner, M. (1993). David Duke and Black Threat: An Old Hypothesis Revisited. *The Journal of Politics, 44*(3), 702–713.

Grady, C. (2019, August 5). The Dayton, Ohio, Shooter Reportedly Kept a 'Rape List' of Potential Victims. *Vox*. Retrieved August 1, 2020, from https://www.vox.com/identities/2019/8/5/20754918/dayton-ohio-shooter-rape-list-mass-shootings-misogyny-link.

Hartzell, S. L. (2020). Whiteness feels good here: interrogating white nationalist rhetoric on Stormfront. *Communication and Critical/Cultural Studies, 17*(2), 129-148. doi:10.1080/14791420.2020.1745858

Hutchinson, A. (2021, January 8). Will the Capitol riots prompt a re-think as to how social platforms address free speech? Retrieved February 20, 2021, from https://www.socialmediatoday.com/news/will-the-capitol-riots-prompt-a-re-think-as-to-how-social-platforms-address/593092/

Johnson, N. F., Leahy, R., Restrepo, N. J., Velasquez, N., Zheng, M., Manrique, P., Wuchty, S. (2019). Hidden resilience and adaptive dynamics of the global online hate ecology. *Nature.* doi:10.1038/s41586-019-1494-7

Maley, P. (2019, September 6). Christchurch shooter 'a role model for extremists.' *The Australian*, p. 6.

Maraniss, D. (1990, July 1). Campaigning in Code. Retrieved May 1, 2020, from https://www.washingtonpost.com/archive/politics/1990/07/01/campaigning-in-code/54d423af-21de-4a98-b5aa-4bb82775b63a/

Meleagrou-Hitchens, A., & Kaderbhai, N. (2017). Research perspectives on Online Radicalisation: A Literature Review 2006-2016. Retrieved May 1, 2020, from http://icsr.info/2017/05/icsr-vox-pol-paper-research-perspectives-online-radicalisation-literature-review-2006-2016/

McArdle, M. (2018, April 10). Sasse Grills Zuckerberg: If You're Going to Police Hate Speech, Can You Define It? *National Review*. Retrieved August 1, 2020, from https://www.nationalreview.com/news/ben-sasse-mark-zuckerberg-hate-speech-policy-definitions/

McBride, J. (2019, December 12). David Anderson Social Media: He Was Likely 'Dawada Maqabath.' *Heavy*. Retrieved June 1, 2020, from https://heavy.com/news/2019/12/david-anderson-social-media-dawada-maqabath/

Morgan, M. (2010). *Mean World Syndrome*. Media Education Foundation. Retrieved June 1, 2020, from https://www.mediaed.org/transcripts/Mean-World-Syndrome-Transcript.pdf.

Müller, K., & Schwarz, C. (2018). *Fanning the Flames of Hate: Hate Crimes and Social Media.* The University of Warwick. Retrieved August 1, 2020, from https://warwick.ac.uk/fac/soc/economics/research/centres/cage/manage/publications/373-2018_schwarz.pdf.

Pane, L.M., & Dazio, S. (2019, September 21). US police assess rise in threat tips after 3 mass killings. *AP News*. Retrieved May 1, 2020, from https://apnews.com/dd054d08eb904d3a92d9c46cb65c489b

Parvini, S. (2019, April 28). Poway synagogue shooting suspect linked to anti-Semitic Internet manifesto. *Los Angeles Times.* Retrieved June 1, 2020, from https://www.latimes.com/local/lanow/la-me-ln-synagogue-shooting-john-earnest-san-diego-20190428-story.html

Perrigo, B. (2019, March 15). 'A Game of Whack-a-Mole.' Why Facebook and Others Are Struggling to Delete Footage of the New Zealand Shooting. *Time Magazine*. Retrieved June 1, 2020, from https://time.com/5552367/new-zealand-shooting-video-facebook-youtube-twitter/

Phillips, J., & Yi, J. (2018). Charlottesville Paradox: The 'Liberalizing' Alt-Right, 'Authoritarian' Left, and Politics of Dialogue. *Society, 55*(3), 221–28.

Pickert, K. (2014). A Murderous Spree. *Time Magazine, 183*(22), 14.

Pohjonen, M. (2019). A Comparative Approach to Social Media Extreme Speech: Online Hate Speech as Media Commentary. *International Journal of Communication, 13*, 3088–3103.

Quinn, M. (2020, March 22). Mark Zuckerberg: 'Optimistic' AI tools to flag, remove hate speech on Facebook will be developed in 5-10 years. *Washington Examiner*. Retrieved May 1, 2020, from https://www.washingtonexaminer.com/news/mark-zuckerberg-optimistic-ai-tools-to-flag-remove-hate-speech-on-facebook-will-be-developed-in-5-10-years

Roose, K (2018, October 28). On Gab, an Extremist-Friendly Site, Pittsburgh Shooting Suspect Aired His Hatred in Full. *New York Times.* Retrieved May 1, 2020, from https://www.nytimes.com/2018/10/28/us/gab-robert-bowers-pittsburgh-synagogue-shootings.html

Sayimer, I., & Derman, M. (2017). Syrian Refugees as Victims of Fear and Danger Discourse in Social Media: A YouTube Analysis," *Global Media Journal: TR Edition 8*(15) 384–403.

Scrivens, R., Gill, P., & Conway, M. (2020). In The Role of the Internet in Facilitating Violent Extremism and Terrorism: Suggestions for Progressing Research. In T.J. Holt & A.M. Bossler (Eds.), *The Palgrave Handbook of International Cybercrime and Cyberdeviance* (pp. 1-20). Cham: Springer International Publishing.

Simon Wiesenthal Center 2019 Digital Terrorism and Hate Report Card (2019, March 14). Retrieved June 1, 2020, from http://www.wiesenthal.com/about/news/2019-digital-report-card.html

Twitter Rules and Policies (n.d.). Twitter. Retrieved February 22, 2021, from https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy.

Von Behr, I., Reding, A., Edwards, C., & Gribbon, L. (2013). *Radicalisation in the digital era. The use of the internet in 15 cases of terrorism and extremism*. Brussels: RAND Europe.

Warner, B. (2010). Segmenting the electorate: The effects of exposure to political extremism online. Communication Studies, 61(4), 430-444. doi:10.1080/10510974.2010.497069

Wojcieszak, M. (2009). "Carrying online participation offline": Mobilization by radical online groups and politically dissimilar offline ties. *Journal of Communication, 59*(3), 564–586.

Youngblood, M. (2020). Extremist ideology as a complex contagion: the spread of far-right radicalization in the United States between 2005 and 2017. *Humanities and Social Sciences Communications, 7*(1), 49. doi:10.1057/s41599-020-00546-3

YouTube Policies (n.d.). YouTube. Retrieved February 22, 2021, from https://support.google.com/youtube/answer/2801939?hl=en

Zannettou, S., Bradlyn, B., De Cristofaro, E., Kwak, H., Sirivianos, M., Stringini, G., & Blackburn, J. (2018). *What is gab: A bastion of free speech or an alt-right echo chamber*. Presented at the Companion Proceedings of the Web Conference 2018.

About the JD Journal for Deradicalization

The JD Journal for Deradicalization is the world's only peer reviewed periodical for the theory and practice of deradicalization with a wide international audience. Named an "essential journal of our times" (Cheryl LaGuardia, Harvard University) the JD's editorial board of expert advisors includes some of the most renowned scholars in the field of deradicalization studies, such as Prof. Dr. John G. Horgan (Georgia State University); Prof. Dr. Tore Bjørgo (Norwegian Police University College); Prof. Dr. Mark Dechesne (Leiden University); Prof. Dr. Cynthia Miller-Idriss (American University Washington); Prof. Dr. Julie Chernov Hwang (Goucher College); Prof. Dr. Marco Lombardi, (Università Cattolica del Sacro Cuore Milano); Dr. Paul Jackson (University of Northampton); Professor Michael Freeden, (University of Nottingham); Professor Hamed El-Sa'id (Manchester Metropolitan University); Prof. Sadeq Rahimi (University of Saskatchewan, Harvard Medical School), Dr. Omar Ashour (University of Exeter), Prof. Neil Ferguson (Liverpool Hope University), Prof. Sarah Marsden (Lancaster University), Prof. Maura Conway (Dublin City University), Dr. Kurt Braddock (Pennsylvania State University), Dr. Michael J. Williams (The Science of P/CVE), and Dr. Aaron Y. Zelin (Washington Institute for Near East Policy), Prof. Dr. Adrian Cherney (University of Queensland).

For more information please see: www.journal-derad.com

Twitter: @JD_JournalDerad
Facebook: www.facebook.com/deradicalisation

The JD Journal for Deradicalization is a proud member of the Directory of Open Access Journals (DOAJ).

ISSN: 2363-9849

Editor in Chief: Daniel Koehler