Self-Radicalization & Extremist Violence

UNIVERSITY of VIRGINIA
FRANK BATTEN SCHOOL of LEADERSHIP and PUBLIC POLICY
BRADY
UNITED AGAINST GUN VIOLENCE

Table of Contents

Lucy R. Fowell

Master of Public Policy Candidate

Frank Batten School of Leadership and Public Policy

University of Virginia

Prepared for the Brady Camping April 2022

Client Overview

This report was prepared for the Brady Campaign, an American non-profit organization that advocates for gun control and against gun violence. The Brady are campaign are committed to research that investigates the trends in gun violence and its growth. Staff from the Brady Campaign have aided in the guidance of this project to understand how self-radicalization is linked to increased gun violence in the United States.

Disclaimer

The author conducted this study as part of the program of professional education at the Frank Batten School of Leadership and Public Policy, University of Virginia. This paper is submitted in partial fulfillment of the course requirements for the Master of Public Policy degree. The judgments and conclusions are solely those of the author, and are not necessarily endorsed by the Batten School, by the University of Virginia, or by any other agency.

Honor Pledge

On my honor as a student, I have neither given nor received unauthorized aid on this assignment.

Lucy Fowell

SECTION 1: INTRODUCTION TO THE POLICY TOPIC

Abstract

Individuals are becoming self-radicalized through the immersion of extremist content available online. Social media platforms play a significant role in facilitating radicalization through their algorithms promoting increasingly extreme content. This project aims to find optimal solutions to help curb self-radicalization and the violence it causes.

Introduction

The creation of the internet has allowed for a highly dynamic means of communication with few barriers to entry. Individuals can communicate with relative anonymity quickly and effectively across borders to an almost limitless audience (United Nations, 2012, p.3). Social media and content-sharing sites have become the most popular platform for these messages. Unlike more traditional forms of communication, these sites allow users to respond and engage.

The growth of the internet has brought about the sinister cultural phenomenon of online radicalization and self-radicalization. The United Nations defines radicalization as indoctrination that often accompanies the transformation of recruits into individuals determined to act with violence based on extremist ideologies (United Nations, 2012, p.6). Online radicalization refers to when this process primarily occurs online. However, this project will focus on self-radicalization. Self-radicalization does not involve users directly associated with a terrorist organization or its leaders. Al-Lami explains that self-radicalization consists of 'individuals becoming familiar with and influenced by radical ideologies without even socializing with radical groups' (2009, p. 7). What distinguishes self-radicalization from online radicalization is that it takes place in isolation, whereby no contact is made directly with terrorist groups, whether in person or virtually (Von Behr et al., 2013 p.20). This project operates with these definitions and conceptual understandings.

Self-radicalization is becoming an increasing area of concern (Hollewell & Longpré, 2021, p.20) and contributed to the indoctrination of the perpetrators of the 2013 Boston Marathon massacre, 2014 Santa Barbara massacre, and the 2015 Charleston church mass shooting (Kydd, 2021; Alfano et al., 2018; Witt, 2020). They were not directly associated with a terrorist organization or leaders but instead consumed various extremist materials and engaged in online groups contributing to their beliefs and violence (Alfano et al., 2018).

The need for greater research into this phenomenon has been galvanized by the extremely violent nature of the January 6th attack. This attack demonstrated the effects of mass-radicalization but more importantly, mass-mobilization, culminating in the largest attack on the U.S. Capitol since the war of 1812 (Holpuch, 2021). The insurrection has brought to light the severe threats of self-radicalization in the landscape of national security threats.

This report hopes to aid in understanding this pressing national security issue and offer policy solutions to counter it. It will first attempt to succinctly define the problem before providing a background of the issue and existing legislation. It will then review academic literature to highlight leading solutions in the field. Finally, it considers each policy to determine an optimal solution moving forward.

Definitions

Defining and redefining concepts is a distinct issue within the field of terrorism and extremism. The slight changes in definitions from author to author do not inherently further discussions within the subject; therefore, an overview of useful concepts used within this project are listed below.

Online Radicalization

The process by which individuals, through interactions with and exposure to various types of internet content and groups, come to adopt beliefs that not only justify violence but compel it to the point where these beliefs translate into violent action (Borum, 2011, p.8)

Self Radicalization

Consists of individuals becoming indoctrinated and influenced by radical ideologies without even socializing with radical groups (2009, p. 7). What distinguishes it from radicalization via the internet is that it takes place in isolation, and implies a process whereby no contact is made with other terrorists or extremists, whether in person or virtually (Hollewell and Longpré, 2021).

Incel

Short for "Involuntarily Celibate," an incel is a member of an online community of young men who consider themselves unable to attract women sexually, typically associated with hostile views toward women and men who are sexually active. Hoffman et al. define an incel-motivated terrorist attack as "the perpetrators publicly expressed their motivation and were primarily driven by their inceldom—that is, their declared affiliation with, and adherence to, the incel movement." (2020, p.569).

Filter Bubble

The way in which a user's future online exposures are decided for them and where certain content will be made more available or even recommended to them based on the algorithms' perceptions of their preferences (Wolfowicz, 2015, p.1).

(Filter bubbles can result in an online "echo chamber")

Echo-Chamber

Echo chambers are networks of like-minded people and are prone to polarization, where their worldview reflects only their radical ideology and simultaneously hinders exposure to potential counter-messaging (Wolfowicz, 2015, p.1).

Problem Statement

Too many Americans are becoming self-radicalized online, resulting in a growing number of extremist attacks.

Increased use of the internet over the last decade has provided a significant change in the process of radicalization. Violent extremist attacks are increasingly motivated by radical beliefs rather than by association with a terrorist group or network. Data from the Global Terrorism Index shows a rise of individual unaffiliated terrorist attacks from under five percent in the mid-1970s to above 70% between 2014 and 2018 (Institute for Economics & Peace, 2019 p.49).

An Intelligence Community report commissioned by President Biden in the wake of the January 6th attack affirmed this threat. It stated that "Domestic Violent Extremists (D.V.E.s) who are motivated by a range of ideologies and galvanized by recent political and societal events in the United States pose an elevated threat to the Homeland in 2021." (Office of the Director of National Intelligence, 2021, p.2). More specifically, they concluded, "D.V.E. 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". The increase in unaffiliated terrorism seen across the United States suggests that attacks will continue to rise without intervention.

SECTION 2: BACKGROUND AND SCOPE

Issue History

Extremist organizations have been able to capitalize from the internet since its inception, surprisingly being one of the earlier adopters of the new technology. Five years before the term World Wide Web was even coined, White Supremacists established the "White Aryan Resistance" online bulletin board (Smith, 2017; Winter et al., 2020, p.5).

During the early years of the internet, some of the key objectives for extremists online were to recruit new individuals, connect with like-minded groups, and promote their image (Gerstenfeld et al. 2003, p. 37-40). Even during this period, extremist sites used sophisticated techniques of persuasion. Extremist groups' numbers, types, and techniques evolved in the following years. It allowed them to: produce and publish propaganda, raise funds and accumulate resources, recruit and mobilize new supporters, and plan and coordinate attacks (Weimann 2004, p. 5-10).

In the last two decades, there has been a substantial growth of persons committing violent extremist attacks that have not been groomed or radicalized by these traditional methods (Institute for Economics & Peace, 2019 p.49). Some of those who go on to commit violent attacks claim to be affiliated with extremist groups, while others cite mixtures of personal grievances and ideological elements as motives (Hamm and Spaaj, 2015 p.7). This notes a shift where individuals can become radicalized without recruitment efforts of terrorist groups and they follow their own personal ideologically extreme beliefs (Von Behr et al., 2013 p.20).

The Department of Homeland Security explains that self-radicalized violent extremists are "motivated by a variety of domestic terrorist ideologies, such as racially- and ethnically-motivated extremism, including white supremacist violent extremism, anti-government, and anti-authority violent extremism, and other ideological strains that drive terrorist violence." (Department of Homeland Security, 2019 p.4) This mixture of ideologies motivated by personal grievances without the active encouragement of a terrorist group is new territory for academic studies.

Self-Radicalization Examples

It is advantageous to explore some prominent categories where the self-radicalization process occurs. Not all members of these groups are self-radicalized; however, those who are self-radicalized often share an interest in the following groups' ideologies. They are also not mutually exclusive; often, self-radicalized individuals will hold beliefs originating from many extremist ideologies.

Inceldom

The "Incel" movement (standing for involuntary celibate) originated in the early 90s as an online space for those struggling with relationships (Hoffman et al., 2020). Over the last three decades, this online-based community has metamorphosed into a breeding ground for deeply misogynistic and violent sentiments (O'Malley, 2020). It is not classed as a terrorist group as it does not have a core universal ideology, leader, core text, or end goal. However, many violent attacks have been linked to users deeply engaged with incel-content.

The first recorded attack attributed to inceldom was in 2014 at the hands of Elliot Rodger (Witt, 2020). Rodger attempted to attack a sorority house at the University of California, Santa Barbara, before targeting random passersby on the streets. He left behind a 133-page manifesto titled My Twisted World, highlighting the need to seek revenge on women who rejected him and made it difficult for him to seek out sexual relationships (BBC, 2018). Rodger was engaged in multiple online incel communities prior to his death and, after the attack, became a hero within the incel community. The phrase "going E.R" was adopted into the group's distinct lexicon to mean committing a violent attack as Rodger did (Cottee, 2020). This phrase appeared as many more have gone on to commit violent attacks in the name of inceldom. There are notable cases such as:

The 2015 Umpqua Community College shooting. Chris Harper-Mercer killed nine people and injured eight before killing himself. Harper also left a manifesto behind, listing his grievances of being a virgin and not having a girlfriend. He also praised killers like Rodgers as "people who stand with the gods" (Percich, A. 2021).

In May 2020, Armando Hernandez Jr. opened fire in Westgate City Center in Glendale, Arizona injuring three members of the public. Hernandez openly admitted to being an incel and stated he had hoped to have killed couples during the attack (Ware, 2021).

Most recently, In August 2021, Jake Davison went on a shooting rampage, killing five people in Plymouth, UK. Davidson had posted in an 'IncelExit' forum looking for a way out of the online culture but admitted his behavior had been permintaley changed (Townsend, 2021). This was Britain's worst mass shooting in ten years and denoted an international growth in incel-related violence.

Right-Wing Extremism

Right-wing content and rhetoric is widespread online, and today individuals can become self-radicalized without having formal links to right-winged terrorist groups or leaders. The popularity of right-wing extremist material has increased in the wake of hyper-polarization, media polarization, and the mainstreaming of conspiracy groups (Kydd, 2021). Although terrorists groups have been the foundation of some of the most prolific violent right-wing attacks, they have also included self-radicalized attackers.

The 'Unite the Right' rally in Charlottesville, Virginia, in 2017 brought together an alt-right network of KKK members, neo-Nazis, militiamen, Trump supporters, and other white supremacists to protest the removal of a Robert E. Lee statue (Blout and Burkart, 2020). This resulted in one of the largest in-person hate-themed meetings in the United States in decades (Johnson, 2018). Self-radicalized individuals and small groups came together with these established groups organized primarily on 'the r/The_Donald,' a Reddit community (Alfon et al., 2018). This is an interesting case as many self-radicalization examples observe "lone-wolf" terrorism. These actors, on the other hand, were connected through their ideological beliefs and mobilized.

Some academics concluded that the radicalization of Trump voters may be less extreme than other forms of radicalization (Alfon et al., 2018, p. 3). Unsurprisingly, literature published after January 2021 is less sympathetic to this view. In similar case to Charlottesville, the Capitol insurrection brought together established terrorist and extremist groups along with a significant number of small groups and individuals that had been self-radicalized. Kydd has dubbed this "mass self-indoctrination" (Kydd, 2021, p.1).

Conspiracy Groups

Many self-radicalized Trump supporters present at the insurrection also intensely follow and believe in conspiracy theories (Uscinski, 2020). Followers of the notorious QAnon conspiracy group were identified as some of those in the capitol invasion (Mangan, 2021). QAnon has a loose Internet-based following numbering in the millions, and its believers claim that a satanic cabal of pedophiles and cannibals controls world governments and the media (Moskalenko and McCauley, 2021). There is now a greater cross-over with religion, with churches explicitly pointing their followers towards extremist online materials. Currently, 1 in 5 white evangelicals now believes in elements of QAnon (Russonello, 2021). Though some influential individuals are active in the movement, it is not an organized group with defined leadership.

Conspiracy theories have increased in popularity, especially in the wake of the COVID-19 pandemic. They have become increasingly mainstream, with 25% of Americans inclined to believe that the outbreak was planned (Schaeffer, 2020). In a time where information was often scarce and sometimes conflicting, this left spaces for individuals to promote their own extreme views to be passed off as news. Individuals with little or no organizational backing could put out information that would travel as far as a news broadcast (Kydd, 2021). These groups and individuals have not been linked to violent acts from their conspiracy theory beliefs alone (although many hold extreme right-wing views discussed above). Yet, they serve as an example of the type of self-radicalization that is possible.

It is important to highlight these sectors, as a recent study revealed that posting on violent right-wing extremist and incel forums increased significantly following the declaration of the pandemic, and the same was not true of left-wing or jihadist forums (Davies et al., 2021).

Self-Radicalization Process

Although literature concerning why more people are becoming self-radicalized is somewhat sparse, academics are far more unified on the question of *how*. Research has pointed to online communities enabling members to quickly become immersed in extreme content for extensive periods. Users then experience the effects of echo chambers, only consuming content and opinions that align with their own, which hardens their beliefs (O'Hara & Stevens, 2015). Social media sites introduce and encourage users toward this content through their algorithms.

Social media sites operate with the personalization of user experience at their forefront. As individuals see content tailored perfectly to their likes and interests, they are more likely to stay on sites, increasing revenues for social media giants (Tworek, 2019). The personalization of a user's experience can be achieved through manually entering information, e.g., hometown, birthday, and hobbies, into a social media profile. Alternatively, data can be gathered through the content users engage with online such as commenting on groups, reacting to political pages, etc. Through both means of information acquisition, algorithms can promote content and advertisements that resonate with users (Agarwal and Sureka, 2015, p. 2-10).

Wolfowicz explains that each time users engages, passively or actively, with content algorithms further refines the information they will later see. Eventually, a user is bound to their "filter bubble" ; each person's bubble is unique, and it presents them with a bias (their own). From this, users can become part of an online "echo chamber," characterized by homogeneity and where the polarization of ideas can lead to the adoption of more extreme stances (Wolfowicz, 2015). Wolfowicz also points out that as users are largely unaware that the information they are consuming is biased, they likely do not believe it is.

In an internal investigation, Facebook created a fake profile to understand this phenomenon more. They created a pretend user - Carol Smith - who was supposedly a politically conservative mother from North Carolina. Smith's indicated interest in politics, parenting, and Christianity. Though Smith had never expressed interest in conspiracy theories, in just two days, Facebook was recommending she join groups dedicated to QAnon. Within one week, Smith's feed was full of groups and pages that had violated Facebook's own rules, including those against hate speech and disinformation (Zadrozny, 2021).

Similarly, a study into YouTube's practices saw that users have a 6.3% chance of being suggested an incel-related video recommendation within five clicks when starting from a non-incel-related video (Papadamou et al., 2020). This is a serious cause for concern as 70% of all YouTube videos watched on the site come via their recommendations to users (Popken, 2021). This aspect of online echo chambers may be referred to as "algorithmic deviancy amplification," where a user engaging with deviant material, content, and associations is fed only increasingly extreme material at an increased frequency, providing stronger reinforcement of these views (Wood, 2016, p.10-13).

International Comparison

This problem is distinctly global. The internet is not bound by land borders which means the issue of online radicalization is similarly spreading in other nations, most notably the UK and Canada. Public attention to this phenomenon was first heightened in the UK after the Manchester Arena bombing in 2017. The perpetrator, Salman Ramadan Abedi, was not a formal member of a terrorist organization, but he consumed extremist materials produced by these groups and acted alone (Herath and Whittaker, 2021). This fear has sustained into 2021 as the nation’s second member of parliament murdered within five years by a self-radicalized individual (Turner, 2021). Canada has followed a similar trend with several incel-related attacks including a 2018 truck attack and a stabbing in early 2020 (Sganga, 2022). In Europe a similar pattern has emerged. There has been a range of ideologically extreme motivated attacks, including White Nationalism, Inceldom, and Jihadist inspired violence (Kupatadze and Argomaniz, 2019; Auger, 2020).

U.S Context

It cannot be understated how universal this issue is across the globe. However, certain political and cultural conditions within the United States have made the outcome of extremist attacks more common and less preventable. This refers to the constitutional and cultural significance of the right to free speech and the right to bear arms.

American citizens value the right to freedom of speech more deeply than other nations, regardless of situational contexts. Over half of Americans support the right to violent speech for the sake of freedom of speech; comparatively, in Asian countries, it's less than a quarter (Suiter & Carolan, 2019). In a study of 38 nations, the U.S. was the nation with the highest percentage of support for the right of others to make statements that are offensive to their own religious beliefs, at 77% (Pew Research Center, 2021).

Being constitutionally and culturally bound by freedom of speech means that current online content regulations are weak. This has allowed individuals within the U.S. to act with relative freedom online. Most significantly, those who have become self-radicalized have unique access to arms to carry out extremist attacks. The Domestic Violent Extremism report commissioned by Biden suggested this reason was why lone offenders pose significant threats to national security (Office of the Director of National Intelligence, 2021).

SECTION 3: POLICY SOULTIONS

Literary Review

Well thought out and data-driven solutions are needed to help curb this ever-growing issue. Literature concerning policy solutions in this domain are generally split into a technological approach and a counter-narrative approach. Put briefly, regulating social media to prevent radical content from being consumed and alternatively putting forward counter-narratives so citizens can independently spot signs of extremism. The technological approach is more widely used, whereas the counter-narrative approach is still in its earlier days of inception.

Technological Approach

Platform Removal

Ribeiro et al.'s thorough study examines the migration of two popular extremist groups on the platform Reddit (Ribeiro et al., 2021). The two groups, "Incels" and "The Donald" were banned from the forum and migrated to their own standalone websites. In both cases, moderation measures significantly decreased posting activity on the new platform, reducing the number of posts, active users, and, most significantly, newcomers. This study presents evidence against the popular theory that regulating extremist groups will lead them to more unregulated spaces and subsequently to become even more extreme than before. This is supported by Jhaver et al.'s study focusing on the effect of Twitter's ban on extremist alt-right influencers Alex Jones, Milo Yiannopoulos, and Owen Benjamin. They found that posts referencing these influencers dropped by an average of nearly 92% within six months of their bans (2021). Additionally, the study found that the influencers' followers who remained on Twitter exhibited a modest but statistically significant drop of about 6% in the "toxicity" levels of their subsequent tweets.

Conversely, Ali et al. found that when individual users were banned (as opposed to groups and celebrities) from Twitter and Reddit, users exhibited an increased level of activity and toxicity on Gab, although the audience they potentially reached decreased.

Reliability and Limitations

Ribeiro et al.'s research design utilized over 6 million posts made by more than 138 thousand users using a robust quantitative method. Most significantly, their conclusions highlighting the reduction in newcomers confirm the direct link between censorship and radicalization prevention. A limitation is that this study only follows two groups. That said, these two communities are among the most prominently sanctioned groups, which provides early insight into the consequences of such sanctions. Jhaver et al. also utilized an effective quantitative method working with over 49 million tweets and measuring content through a toxicity score. Although its findings disagrees with the overarching concept of censorship, Jhaver et al. have found a unique gap in the literature that provides a more nuanced picture of censorship. It has different effects on different audiences that not all studies have been able to capture.

Algorithms

Whittaker and Herath's study monitoring the effects of algorithms when interacting with far right-wing content found that Youtube amplifies extreme and fringe content, while two Reddit and Gab do not. Another study on Youtube's role in promoting extremist content found that 9.4% of recommended videos would be incel-related if a user has seen two incel-related videos and 11.4% if a user had seen three incel-related videos. Moreover, the portion of Incel-related recommendations increases substantially as the user watches an increasing number of consecutive incel-related videos (Papadamou et al., 2020). Massanari's case study found similar results and exposed how Reddit's algorithms, policies, and general community structure enables and even support extremist content (Massanari, 2017). Discussing policy implications, Whittaker and Herath state current solutions to algorithm amplification focus on transparency rather than "de-amplification." They argue knowing understanding the effects of algorithms is not enough to prevent radicalization.

Reliability and Limitations

Papadamou et al. 2020 and Massanari 2017 used different approaches in their studies. Massanari uses two case studies of what is defined as "toxic technocultures" in a discussion piece evaluating Reddit's policies. This is a purely qualitative piece which, of course, creates its own set of limitations. That said, overall, the findings were still reliable, citing a range of credible sources and showing thorough investigative work. On the other hand, Papadamou et al. used a far more quantitative approach collecting over 6500 YouTube videos shared on incel-related forums and building a lexicon of 200 incel-related terms. This was finally cross-referenced with video content to determine if it was incel in nature. There was a notable limitation with this method; sometimes, a false positive could occur in which a non-incel-related video was flagged as incel in nature. However, this was a minimal number in the final results after using countermeasures. This did not compromise the reliability of these findings. Whittaker and Herath also offered a highly quantitative piece that uses automated user accounts to observe content adjustments on social media sites. Using Youtube and Reddit in this investigation is appropriate as they are mainstream and used by a range of users regardless of their political or ideological beliefs. The examination of the social media site Gab, however, is a much weaker comparison. Gab is a social media platform primarily used by conservatives and right-wing extremists. Those who have already become self-radicalized use communication channels such as this, so understanding their recommendations is less fruitful. Despite this, the study's results on Youtube and Reddit behavior are helpful to the wider academic understanding of algorithms' role in self-radicalization.

Content Removal

Milton’s 2016 study shows the self-regulation of social media sites, especially Twitter, reduces the quantity of extremist content online. He finds posts on anonymous distribution accounts on Twitter are being removed in greater numbers and quicker than ever before (Formica, 2020). These findings are consistent with others in the field. Berger and Morgan's study also focuses on Twitter and found the amount of pro-ISIS content available was significantly reduced by account suspensions as all of a user's tweets are deleted when their account is deleted (Berger & Morgan, 2015). While valuable, Milton's 2016 and Berger and Morgan's 2015 studies are only focused on Twitter and ISIS-related content. More comprehensive studies find similar conclusions, but it is worthy to note they only look at a subsect of the issue.

Looking into content removal technology, Van Der Vegt et al. point to the success of automated detection infrastructure in targeting material related to violent white supremacist content (2019). They argue an automated terrorist content detection system that can flag suspect messages or accounts for humans to review offers a effective way to limit sharing. They believe a hybrid of human and machine removal is needed for the most successful approach.

Reliability and Limitations

Berger and Morgan's content analysis method only uses 14 declassified documents; a very small sample of work casts doubts on the reliability of their results. Van Der Vegt et al. has been able to produce a more comprehensive overview; however, they fail to fully consider the ability of social media sites to produce hybrid approach models of content removal.

Counter-Narrative Approach

A growing body of literature explores the need for counter-communication campaigns to offset extremist materials. In total, to understand the allure of extremist videos and the ineffectiveness of U.S. video messages designed to 'counter violent extremism (Bean et al., 2017)

Weimann and von Knop theorize solutions are embedded in a more sophisticated strategy than censorship, promoting the theoretical notion of “noise reduction” in countering terrorist communications (Weimann & von Knop 2008). This is also supported by Bean and Edgar's conclusions, which state that if the adversary's message is inherently more appealing because of its construct skill, attempts at countering it are bound to fail (Weimann & von Knop 2008). Neumann and Stevens argue that any strategy that relies on reducing the availability of content alone is bound to be “crude, expensive and counterproductive” (Stevens & Neumann, 2012). McDowell-Smith et al. go further by testing how U.S students react to carefully curated terrorist defector interview videos. They found students were likely to turn them away from ISIS (McDowell-Smith et al., 2017).

Archetti argues against McDowell-Smith et al. They offer that it is not enough to simply will individuals to act in a given way. They suggest that if the study were repeated with focus groups containing young people from northern Iraq, it would undoubtedly return different results (Archetti, 2018). This, however, is a redundant analysis when in search of policy solutions to prevent the radicalization of U.S citizens. Despite this counter, overall, there is a consensus within scholarship that utilizing other solutions outside of the technology approach can only aid in efforts to prevent radicalization.

Reliability and Limitations

Utilizing a relatively unique method of sound analysis, Bean and Edgar go into great detail about how U.S counter-terrorism videos require higher quality and allure to begin to match the appeal of extremist content. Their methodology is robust; however, their study does more to point out problems within current methods rather than state the impact if such changes are made. This is a similar issue found in Weimann and von Knop's piece, yet this was further limited due to its weaker research design. Weimann and von Knop systematically evaluated the types of "noises" institutions could make to drown out extremist communications. This relied heavily on anecdotal evidence rather than a quantitatively based method.

Stevens and Neumann do a good job reviewing each policy option currently in operation and evaluating its effectiveness. For example, it shows how content filtering is an expensive tool that can block permitted websites. Whereas search engine optimization, which promotes safer sites above more dangerous ones, is widely frowned upon and may have limited success. This report is limited as it fails to mention the data used to reach these conclusions, so its reliability is difficult to comment on.

Mcdowell et al. are the only study to use a focus group approach. They produced two video clips of ISIS defectors denouncing the group, which were focus tested by a college student sample of 75 undergraduate students. These results showed that 90% responded these videos would help prevent others from joining ISIS; this study result is encouraging but greatly limited. The sample of 75 college students is again small, so generalizations about results should be restrained. Additionally, the results themselves tell us very little about whether self-radicalization can be prevented through counter-narrative videos. It only found consumers thought it could prevent *others* from joining ISIS. While this study showed the most traditional research design, it failed to ask the right questions.

Literary Review Reflection

Technological and counter-narrative approaches have their merits; however, there is overwhelming literature supporting the need for tech-based solutions. As social media platforms contribute so greatly to this problem, moderate reforms could have a significant impact. On the other hand, a counter-narrative approach is still a particularly niche field that requires more sound research to fully appreciate its potential benefits.

Existing Laws

Alongside academic evaluation, consideration of existing laws and policy jurisdiction is imperative for recommending solutions. Legislation in this domain can be traced back to the Communications Decency Act of 1996. Specifically, Section 230 states, "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). This provision ensures social media sites face no responsibility for their users' behaviors. Further, social media platforms can't be sued for taking down or leaving content up (Reardon, 2021). Its simple language removes the common law's distinction between publisher and distributor liability, granting operators of social media sites broad protection from claims based on the speech of third parties (Ardia, 2009, p.1). In the 25 years since its enactment, section 230 remains one of the most significant statutes concerning social media liability. Many commentators have suggested Section 230 has allowed internet firms to avoid accountability for the malfeasance that their platforms have enabled. They also argue that Section 230 provides social media platforms a permanent excuse to avoid any discussion of their ethical responsibilities (Cramer, 2020; Goldman, 2018; Spiccia, 2013).

Issues with Section 230

Democrats have advocated for social media companies to take down hate speech, harassment, disinformation, and terrorism-related content. They have even accused the companies of using the liability protections to profit from the 'lies' spread on their platforms (Reardon, 2021). Conversely, Republicans have alleged that social media companies censor conservative viewpoints, with former President Trump taking particular issue with this during the 2020 election. Although many in day-to-day life acknowledge social media platforms as one of the most significant platforms for speech, it is not public at all. Such "public forums" are governed by private firms' policies and procedures (Ardia, 2009). Campaigns for change produced by the Republican party will likely center around aligning private social speech and free speech protected by the constitution.

No matter which side of the aisle reform comes from, it is likely that section 230 will see alteration. This process is expected to be expedited in the wake of the 2021 Capitol attack. President Biden stated that Section 230 should be "revoked, immediately," and Senator Lindsey Graham (R-SC) has said, "Section 230 as it exists today has got to give." (Reardon, 2021). In the past year Sens. Brian Schatz and John Thune have unveiled the PACT Act, which would force companies to be more transparent about content moderation. In a similar move, a group of Democrats in the House and Senate introduced the Safe Tech Act, which aims to hold tech companies more accountable when posts on their services result in real-world harm (Vynck et al., 2021).

Tech Firms Positions

Facebook CEO Mark Zuckerberg has expressed an openness to changing Section 230. In contrast, Google CEO Sundar Pichai and Twitter CEO Jack Dorsey have defended the act (Vynck et al., 2021). During a congressional hearing earlier this year, Pichai said he has concerns about changing or repealing the law, noting potential consequences such as harming free expression (Mills-Rodrigo, 2021). Dorsey echoed Pichai's concerns that restrictions could be difficult to enforce and could have unintended consequences, especially for smaller platforms (Reardon, 2021).

Policy jurisdiction

Executive Commissions

The Federal Communications Commission regulates interstate and international communications through cable, radio, television, satellite, and wire. The goal of the Commission is to promote connectivity and ensure a robust and competitive market.

The Federal Communications Commission's mandate covers the cultivation of a robust and equitable communications infrastructure; however, they do have some autonomy in interpreting section 230 (Ardia, 2009). The Supreme Court has twice considered whether the FCC’s general rulemaking authority extends to the 1996 amendments to the Act. Both times, the Court held that it does (Cheah, 2020). The FCC has been hesitant to act in this area despite its legal authority granted by the courts. The chairman of the FCC under the Trump administration had plans to work on his interpretation of section 230; however, the change in administration prevented anything actionable before leaving the post. New chairwoman Jessica Rosenworcel has stated she did not favor commission action on Section 230, so the status quo will likely prevail from the FCC (Krishan, 2021).

The Federal Trade Commission aims to protect consumers and competition by preventing anticompetitive, deceptive, and unfair business practices through law enforcement, advocacy, and education without unduly burdening legitimate business activity.
The FTC acts as a watchdog over business practices at large, so it is responsible for social media sites as profit-making firms. It cannot regulate social media companies' content; however, it does monitor how it uses citizens' data and information (Coldewey, 2020). In 2020 the FTC ordered nine social media and video streaming companies to provide data on how they collect, use, and present personal information and how their practices affect children and teens (Federal Trade Commission, 2020). This indicates a high level of awareness of the harm of data collection and algorithm practices.

The Federal Communications Commission and the Federal Trade Commission hold some regulatory authority under the executive branch, providing necessary guardrails for the social media industry. That said, there is no federal commission or agency that is directly responsible or capable of dealing with the intricacies of social media firms.

The President

The President has the authority to sign executive orders; however, executive orders can be overturned by the following administration or overridden if they conflict with existing statutes. President Trump signed Executive order 13925 attempting to limit media platforms' alleged censorship of conservative views (Goldman and Miers, 2021). It failed to change anything substantially before Biden revoked it the following year. For the brief time it was in effect It acted more as a request for executive agencies to align their policy priorities with the administration. Even with executive authority Trump was confined to current regulations under section 230 and future presidential action.

State Regulators

States also possess the authority to regulate in their region, which can be effective with enough consensus. In 2018 The California state legislature introduced The California Consumer Privacy Act (CCPA), making social media companies more transparent about their data collection practices (Baik, 2020). In a similar effort, the Illinois state legislature introduced the Biometric Information Privacy Act (BIPA).This limits companies’ collection and use of biometric data like fingerprints and facial recognition. It has been an obstacle for Facebook, Microsoft, and others who have taken for granted the ability to analyze a user’s biological metrics (Buresh, 2021).

Congress

Congressional gridlock is no exception to this policy area. As a particularly divisive topic, it is unlikely Congress would be able to pass any effective legislation in this area. While they have the authority to do so, this is not a practical avenue. That said, this does not negate the fact representatives are paying far closer attention to the effects of social media's role in radicalization. January 6th played a crucial part in highlighting this issue. The House Select committee investigated the insurrection calling social media and tech firm CEOs to answer its questions (Gregorian and Caldwell, 2021). Overall, policymakers are far more aware of the need to legislate, but partisan divides remain.

SECTION 4: POLICY EVALUATION

Policy Alternatives

With consideration of academic literature, existing laws, and known cases of self-radicalized individuals, there are a number of policy solutions worthy of evaluation for this project. Prominent solutions address both the extremist content available to users and further social media firms' promotion of extreme content. There are three credible solutions to address these areas of greatest concern. These are Deplatforming, Algorithm Reform, and Automated Content Removal.

1. De-platforming

De-platforming is the term used to describe the removal of persons or groups from platforms that post extremist materials, hate speech, or violate user agreements. A comprehensive de-platforming strategy would delete accounts of users that post extremist and terrorist-related materials. This would do so irrespective of celebrity or political status. An update of user agreements would be required to introduce specific amendments related to extremist materials, specifically those that encourage violence.

2. Algorithm reform

Algorithm reform focuses on altering the mechanism that recommends increasingly extreme content to users. Effective algorithm reform would place blanket bans on extremist content and significantly alter recommendations of milder content. It would also offer counter content to those users explicitly searching for violent or extreme content. In the long term, algorithms should be developed to learn patterns of users seeking out extreme videos in order to further policy research.

3. Promotion of Artificial Intelligence and Auto-Mated Removal

Automated content removal aims to mass remove content that violates user agreements and policies. Automated removal technology can process vast quantities of data and discover patterns and correlations in the data unseen to the human eye, enhancing the effectiveness of stopping extremist content from being posted. Preventing content from being published, as opposed to requiring removal once live, takes away the risk of content being shared or saved before it can be taken down.

Criteria for Evaluation

This analysis will evaluate potential policy alternatives identified as prominent solutions in the field; It will be measured against criteria tailored to the client. This criterion includes the effectiveness of the policy in preventing self-radicalization and related violence. Further, it considers the political and corporate feasibility of the alternatives. Traditionally policy research hones in on political and legal feasibility; however, policy changes in this domain tend to stem from private firms changing their own user policies. As such, corporate feasibility requires equal consideration in this analysis. Using an outcomes matrix with a ranking system of High-Medium-Low will offer the optimal solution moving forward, given the available data.

Effectiveness (Self-Radicalization) – This measures the estimated likelihood of reduction of those who have been self-radicalized.

Effectiveness (Violence) - This measures the estimated likelihood of reduction of violence by those who have been self-radicalized.

Corporate Feasibility - This measures the estimated likelihood of a program being implemented and sustained by corporations. These factors include public opinions, NGO pressure, and government championing.

Political Feasibility – This measures the likelihood of a political change. These factors include public opinion, NGO pressure, and corporation lobbying.

1. **De-platforming**

Effectiveness (Radicalization) - High
The literature discussed already and other works consistently point towards the effectiveness of de-platforming policies. Yokotani and Takano's investigation into users bans and suspensions found that all suspensions, including 24-h, 48-h, and permanent suspensions, effectively decreased future offenses and damage by suspended offenders and their peers (2022). The reduction of offensive posting by individuals and their peer group after bans follows a similar trend as celebrities and their followers. Jhaver et al.'s study follows the effect of de-platforming of influential users/celebrities and found posts referencing each influencer declined significantly, by 91.77% on average. Additionally, the number of new users tweeting about each influencer also diminished by 89.65% on average (2021).

Effectiveness (Violence) - Low
Some studies found that users who migrated to other platforms often became more radicalized in their new communities. Followers who exhibited more toxic behavior moved to alternative platforms like 4Chan and Gab, which have laxer rules against harmful speech than major social media networks (Ghaffary, 2022). Alternatively, a study into banned Reddit groups who migrated to standalone websites found significantly decreased posting activity on the new platform reduced the number of posts, active users, and newcomers (Ribeiro et al., 2021). Cassie Miller, a senior research analyst at the Southern Poverty Law Center, has commented, "We know that de-platforming works, but we have to accept that there's no silver

bullet; tech companies and governments are going to have to continually adapt." De-platforming has the potential to reduce the number and the level of commitment of radicalized users if approached correctly.

Corporate Feasibility - Medium
After the January 6th attacks, a number of social media platforms significantly re-evaluated their policies regarding extremist and violent content. The events, which were largely planned and promoted using social media, highlighted the role of tech platforms in facilitating such violence. Most notably, this saw the removal of Donald Trump and his content in some form on all major social media platforms, including Twitter, Facebook, and Youtube. As there are existing cases of social media sites changing their policies, we may likely see further changes, and such organizations would be willing to change their existing policies. That said, as significant changes follow from a transformative event like the January 6th attacks, it may take another such event to spark further change.

Political Feasibility - Low
Former President Trump and a number of Republican representatives have taken issue with the ability of social media sites to remove platforms (Gold and Treene, 2022). Conversely, attempts to ensure such removal by more left-leaning representatives holds little legislative ground. Both democrat and republican members find issues with social media sites' ability to remove and keep users with little oversight. Congressional research papers show that government action regulating internet content in this way may implicate the First Amendment. In particular, social media providers may argue that government regulations impermissibly infringe on the providers' own constitutional free speech rights. Legal commentators have argued that when social media platforms decide whether and how to post users' content, these publication decisions are themselves protected under the First Amendment (Brannon, 2019).

2. **Algorithm reform**

Effectiveness (Radicalization) - High
A number of studies have directly linked the role of algorithms in the self-radicalization process (Papadamou et al., 2020); Youtube is one the worst offenders for this consistently recommending extremist-related content. Facebook has found similar issues with its algorithms recommending content that violates its own code of conduct (Zadrozny, 2021). TikTok's recommendation algorithm likewise promotes content from QAnon and far right-wing groups. In one example, after users interacted with transphobic videos on TikTok, the recommendation algorithm fed these users' videos with hate symbols, white supremacist, and anti-Semitic content (Little, 2021). An internal Facebook memo admitted that core product mechanics, including recommendations and optimizing for engagement, are key to why hate and misinformation flourish on the platform (Danner, 2021). With effective reform of algorithms, users would be less likely to find themselves in the rabbit role of extremist content that platforms promote.

Effectiveness (Violence) - Uncertain
Despite the growing attention to the issue, there is little research on algorithm regulation and how it specifically links to violence reduction. Therefore greater research is needed before categorizing its effect.

Corporate Feasibility - Medium
Social media platforms are becoming increasingly aware of their influence on their users. The conduct of internal investigations into the issue demonstrates a certain level of acknowledgment of the issue. Facebook's Vice President for global affairs has stated the company is open to greater transparency; however, Facebook and smaller platforms have taken few real actions towards this (Zorthian, 2021). More importantly, Herath and Whittaker have stated transparency of algorithms is not enough to mitigate self-radicalization (2021). Overall social media firms are slowly adjusting their stance, but their speed does not reflect the seriousness of the threat.

Political Feasibility - Medium
New international norms in this area may prove to influence US policymaking. The Christchurch terror attack in 2019 highlighted the harm of algorithms leading the New Zealand Prime Minister and French President to bring together several heads of state and tech companies to propose the Christchurch Call. The Call committed governments to, amongst other things, 'review the operation of algorithms and other processes that may drive users towards and/or amplify terrorist and violent extremist content' (Christchurch Call, 2019). The Call was signed by the European Commission, the Council of Europe, and 49 nation-states. This growing commitment to algorithms regulation internationally means the US could be likely to follow its allies. Legislators have shown increasing interest in passing a bill that would hold tech firms more accountable for the content it amplifies using its algorithm. But legal experts are divided over whether such a change could survive a legal challenge on First Amendment grounds (Zorthian, 2021).

3. Promotion of Artificial Intelligence and Auto-Mated Removal

Effectiveness (Radicalization) - High
Tech firms have reported success using artificial intelligence to detect extremist content. Facebook in 2017 has stated it's able to remove 99% of Islamic State and Al Qaeda terrorist content before it's flagged by users, thanks to advances in artificial intelligence (Guynn, 2017). Similarly, in 2017, YouTube removed over 150,000 videos for violent extremism, with 98% of these flagged by automated technology (Macdonald, 2019, p.184). The mass removal of content aids in the prevention of self-radicalization, especially when before content can be published.

Effectiveness (Violence) - High
This alternative may also help specifically in the prevention of violence, as AI can be utilized by law enforcement to help predict radicalization and violence before it happens. The

International Center for the Study of Radicalization at King's College London reported a growth in technology that can predict the escalation of online behaviors, allowing relevant authorities to act before violence occurs (Schroeter, 2020).

Corporate Feasibility - High
Large platforms are actively promoting this type of advancement. Not only does it help address the ethical implications of hosting extremist content on its sites, but it also cuts down on the substantial amount of hours required by workers to review potentially harmful content manually. There is a high chance this will be maintained and progress further as already in practice.

Political Feasibility - Medium
The European Union has been consistently reviewing policies that harness this type of technology to limit the spread of extremism internationally. Again, as we see other western democracies adopt these policies, it may influence US policymakers (Elkin-Koren 2020). Alternatively, the increased use of this technology by law enforcement monitoring suspected social media accounts may prove to be an attractive option without the need for private firms.

Outcomes Matrix

	Effectiveness (R)	Effectiveness (V)	Corporate Feasibility	Political Feasibility
De-platforming	High	Low	Med	Low
Algorithm Reform	High	Uncertain	Med	Med
Promotion of Auto-Mated Removal	High	High	High	Med

Recommendation: The outcomes matrix offers that AI technology prompting the automatic removal of content to be the most optimal choice moving forward.

SECTION 5: CONCLUSION

Implementation

Recent figures suggest that 350,000 tweets are posted every minute, 300 hours of video are uploaded to YouTube, and, on Facebook, 510,000 comments are posted, 293,000 statuses are updated, and 136,000 photos are uploaded (Macdonald, 2019, p.184). Given the sheer volume of content, social media platforms already utilize automated removal technology for various extreme materials posted on their sites. This ranges from sexual exploitation to hate speech and spam.

Despite some successes in automated removal, the problems presented in this project remain. Systems of removal largely remain opaque, unaccountable, and poorly understood (Gorwa et al., 2020). Greater resources need to be dedicated to the innovation of automated removal across platforms to sustain efforts against extremism and radicalization. This must be tied in with greater public knowledge of this process to ensure accountability of these platforms' efforts.

Advocacy

Advocacy and education are at the core of The Brady campaign's mission; using these skills, Brady should shift some of its focus from congressional leaders toward influential social media platforms. Brady should focus on the following elements to lobby effectively for greater AI incorporation.

Coalition Partners
Brady has experience working with a variety of liberal NGOs to pursue its goals. This case offers an opportunity to work with new groups more closely linked to preventing extremist content, misinformation, and other advocacy groups within this domain. Organizations such as Color of Change, the Anti-Defamation League, and Common Sense Media are increasingly pushing tech companies to take more aggressive steps to moderate their sites for misinformation, voter suppression, and discrimination. These groups would provide excellent partners, and some have already partnered with Brady on other projects.

Education
Reports such as this serve as an educational tool that aid in Brady's existing efforts to counter misinformation and highlight societal areas in need of reform. Tailoring this report and its findings to publish alongside existing reports conducted by Brady can help put a spotlight on the issue. As Brady is a trusted scour of information, they are a credible basis for citation and further exposure.

Direct Contact

Attempting to set up discussions with social media firms mentioned in this report could open dialogue and sustain relationships. Presenting a willingness to partner and discuss grievances within a mutually respectful environment could aid in both parties achieving their goals. Framing these firms as partners is vital to ensuring alterations to their strategies.

Obstacles

The recommendation in this report of automated content removal is a long-term solution to preventing self-radicalization and gun violence. This is unlikely to reduce gun violence immediately; therefore, it may see some organizational resistance. Further, it may be difficult to directly consult with industry leaders regarding this matter, despite Brady's stature in the NGO space.

Concluding Remarks

This rise of the internet has shaped the way in which we communicate; however, its unintended side effects have left some users extremely isolated and open to persuasive messaging. With information at our fingertips, anyone can slip down the internet rabbit hole, altering our perceptions and understanding of the world. This can lead to a false sense of reality, pushing individuals to commit heinous crimes. As these attacks persist, policy evaluations like this are needed to find feasible and practical solutions. In an attempt to offer solutions, this report settled on the promotion of artificial technology and automated removal of extremist content online. This policy aims to mass remove extremist material that may immerse users in radical territory. While there may be some limitations in The Brady Campaign's ability to influence social media platforms' technology policies, this project provides a broader understanding of self-radicalization.

Key Takeaways

- Self-Radicalized individuals are present across the ideological spectrum, but they do not behave in the same way as traditional terrorist group members. They are motivated through their own mixture of ideologies and beliefs.
- Social media platforms play a significant role in the self-radicalization process through the promotion of radical content and infrastructure that cause echo chambers.
- This is a growing issue internationally; however, it remains a major national security threat due to access to arms.

Offering policy solutions that fit within a complex ecosystem of actors will remain challenging for years to come. Change may only come at the hands of private firms, however, with enough pressure, this may be enough to see considerable improvements.

References

Agarwal, S., & Sureka, A. (2015, February). Using knn and svm based one-class classifier for detecting online radicalization on twitter. In International Conference on Distributed Computing and Internet Technology (pp. 431-442). Springer, Cham

Alfano, M., Carter, J. A., & Cheong, M. (2018). Technological seduction and self-radicalization. Journal of the American Philosophical Association, 4(3), 298-322.

Aly, A., Macdonald, S., Jarvis, L., & Chen, T. (2016). Introduction to the Special Issue: Terrorist Online Propaganda and Radicalization. Studies In Conflict & Terrorism, 40(1), 1-9. https://doi.org/10.1080/1057610x.2016.1157402

Amnesty International. (2021). Retrieved 5 December 2021, from https://www.amnesty.org/en/latest/news/2019/07/the-great-hack-facebook-cambridge-analytica/.

Amnesty International 'The Great Hack': Cambridge Analytica is just the tip of the iceberg. Amnesty International. (2021). Retrieved 5 December 2021, from https://www.amnesty.org/en/latest/news/2019/07/the-great-hack-facebook-cambridge-analytica/.

Archetti, Christina. 2018. The Unbearable Thinness of Strategic Communication. In Countering Online Propaganda and Violent Extremism: The Dark Side of Digital Diplomacy, ed. Comeliu Bjola and James Pamment, 81–96. Abingdon-on-Thames: Routledge.

Ardia, D. S. (2009). Free speech savior or shield for scoundrels: an empirical study of intermediary immunity under Section 230 of the Communications Decency Act. Loy. LAL Rev., 43, 373.

Artificial Intelligence and Countering Violent Extremism: A Primer: Marie Schroeter, GNET, October 2020.

Auger, V. A. (2020). Right-wing terror. Perspectives on Terrorism, 14(3), 87-97.

Baik, J. S. (2020). Data privacy against innovation or against discrimination?: The case of the California Consumer Privacy Act (CCPA). Telematics and Informatics, 52.

BBC News (2018) Elliot Rodger: How misogynist killer became "incel hero." Author. https://www.bbc.com/news/world-us-canada-43892189

Bean, Hamilton, and Amanda Edgar. 2017. A Genosonic Analysis of ISIL and US Counter-Extremism Video Messages. Media, War, and Conflict 10 (3): 327–44

Berger, J. M., and Jonathon Morgan. 2015. The ISIS Twitter Census: Defining and Describing the Population of ISIS Supporters on Twitter. Washington, DC: Brookings

Institution. https://www.brookings.edu/wp-content/uploads/ 2016/06/isis_twitter_census_berger_morgan.pdf.

Blout, E., & Burkart, P. (2020). White Supremacist Terrorism in Charlottesville: Reconstructing 'Unite the Right'. Studies in Conflict & Terrorism, 1-22.

Borum, R. (2011). Radicalization into violent extremism I: A review of social science theories. Journal of Strategic Security, 4(4), 7–36. https://doi.org/10.5038/1944-0472.4.4.1

Brannon, V. C. (2019). Free speech and the regulation of social media content. Congressional Research Service, 45650, 1-43.

Bright, J. (2017). Explaining the emergence of echo chambers on social media: the role of ideology and extremism. Available at SSRN 2839728.

Buresh, D. L. (2021). Should Personal Information and Biometric Data Be Protected under a Comprehensive Federal Privacy Statute That Uses the California Consumer Privacy Act and the Illinois Biometric Information Privacy Act as Model Laws?. Santa Clara High Tech. LJ, 38, 39.

Cheah, M. A. (2020). Section 230 and the Twitter Presidency. Nw. UL Rev. Online, 115, 192.

Coldewey, D. (2020). Who regulates social media. Retrieved 7 April 2022, from https://techcrunch.com/2020/10/19/who-regulates-social-media/

Cramer, B. W. (2020). From Liability to Accountability: The Ethics of Citing Section 230 to Avoid the Obligations of Running a Social Media Platform. Journal of Information Policy, 10(1), 123-150.

Danner, C. (2021). What Is Being Leaked in the Facebook Papers?. Retrieved 7 April 2022, from https://nymag.com/intelligencer/2021/10/what-was-leaked-in-the-facebook-papers.html

Davies, G., Wu, E., & Frank, R. (2021). A Witch's Brew of Grievances: The Potential Effects of COVID-19 on Radicalization to Violent Extremism. Studies in Conflict & Terrorism, 1-24.

Department of Homeland Security. (2019). DHS strategic framework for countering terrorism and targeted violence.

Elkin-Koren, N. (2020). Contesting algorithms: Restoring the public interest in content filtering by artificial intelligence. Big Data & Society, 7(2), 2053951720932296.

Elmas, M. S. (2021). Perceived risk of terrorism, indirect victimization, and individual-level determinants of fear of terrorism. Security Journal, 34(3), 498-524.

Fioretti, Julia. 2017. Social Media Giants Step Up to Join Fight Against Extremist Content. Reuters, 26 June. https://www.reuters.com/article/us-internet-extremism/ social-media-giants-step-up-joint-fight-against-extremist-content-idUSKBN19H20A

Formica, T. (2020). A Social (Media) Contract: Reconciling American Freedom and Security in an Age of Online Radicalization and Extremism. Yale J. Int'l Aff., 15, 131.

FTC Issues Orders to Nine Social Media and Video Streaming Services Seeking Data About How They Collect, Use, and Present Information. (2020). Retrieved 7 April 2022, from https://www.ftc.gov/news-events/news/press-releases/2020/12/ftc-issues-orders-nine-social-media-video-streaming-services-seeking-data-about-how-they-collect-use

Gerstenfeld, P. B., Grant, D. R., & Chiang, C. P. (2003). Hate online: A content analysis of extremist Internet sites. Analyses of social issues and public policy, 3(1), 29-44.

Ghaffary, S. (2022). Does banning extremists online work? It depends. Retrieved 7 April 2022, from https://www.vox.com/recode/22913046/deplatforming-extremists-ban-qanon-proud-boys-boogaloo-oathkeepers-three-percenters-trump

Ghosh, D., & Scott, B. (2021). Facebook's New Controversy Shows How Easily Online Political Ads Can Manipulate You. Time. Retrieved 5 December 2021, from https://time.com/5197255/facebook-cambridge-analytica-donald-trump-ads-data/.

Gold, A., & Treene, A. (2022). Conservatives plot to punish the tech industry for deplatforming Trump. Retrieved 8 April 2022, from https://www.axios.com/republicans-tech-trump-5a85a2dc-8360-4d29-87b1-1da7cc9abfed.html

Goldman, E., & Miers, J. (2021). Online Account Terminations/Content Removals and the Benefits of Internet Services Enforcing Their House Rules.

Goldman, E. (2018). An overview of the United States' section 230 internet immunity.

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. Big Data & Society, 7(1), 2053951719897945

Gregorian, D., & Caldwell, L. (2021). Jan. 6 committee subpoenas tech giants after 'inadequate responses'. Retrieved 7 April 2022, from https://www.nbcnews.com/politics/congress/jan-6-committee-subpoenas-tech-giants-after-inadequate-response-n1287442

Guynn, J. (2017). Retrieved 8 April 2022, from https://www.usatoday.com/story/tech/2017/11/28/facebook-says-artificial-intelligence-has-sped-up-removal-terrorist-content/903615001/

Heeks, M., Reed, S., Tafsiri, M. and Prince, S., 2018. The economic and social costs of crime second edition. Home Office Research report99

Herath, C., & Whittaker, J. (2021). Online Radicalisation: Moving beyond a Simple Dichotomy. Terrorism and Political Violence, 1-22.

Hoffman, B., Ware, J., & Shapiro, E. (2020). Assessing the threat of incel violence. Studies in Conflict & Terrorism, 43(7), 565-587.

Hollewell, G. F., & Longpré, N. (2021). Radicalization in the Social Media Era: Understanding the Relationship between Self-Radicalization and the Internet. International journal of offender therapy and comparative criminology, 0306624X211028771.

Holpuch, A. (2021). U.S. Capitol's last breach was more than 200 years ago. the Guardian. Retrieved 4 December 2021, from https://www.theguardian.com/us-news/2021/jan/06/us-capitol-building-washington-history-breach.

Horta Ribeiro, M., Jhaver, S., Zannettou, S., Blackburn, J., Stringhini, G., De Cristofaro, E., & West, R. (2021). Do Platform Migrations Compromise Content Moderation? Evidence from r/The_Donald and r/Incels. Proceedings of the A.C.M. on Human-Computer Interaction, 5(CSCW2), 1-24.

Institute for Economics & Peace. Global Terrorism Index 2019: Measuring the Impact of Terrorism, Sydney, November 2019. Available from: http://visionofhumanity.org/reports

Jensen, M., James, P., Lafree, G., Safer-Lichtenstein, A., & Yates, E. (2018). The use of social media by United States extremists. START, CollegePark

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

Kastoryano, R. (2017). Radicalization in Europe.

Kupatadze, A., & Argomaniz, J. (2019). Introduction to Special Issue–Understanding and conceptualizing European jihadists: Criminals, extremists or both?. European Journal of Criminology, 16(3), 261-277.

Kydd, A. H. (2021). Decline, radicalization and the attack on the U.S. Capitol. Violence: An International Journal, 26330024211010043.

Little, O. (2021). TikTok is prompting users to follow far-right extremist accounts. Retrieved 7 April 2022, from https://www.mediamatters.org/tiktok/tiktok-prompting-users-follow-far-right-extremist-accounts

Lomas, N. (2021). TechCrunch is part of the Yahoo family of brands. Techcrunch.com. Retrieved 6 December 2021, from https://techcrunch.com/2021/04/29/eu-adopts-rules-on-one-hour-takedowns-for-terrorist-content/.

Macdonald, S., Correia, S. G., & Watkin, A. L. (2019). Regulating terrorist content on social media: automation and the rule of law. International Journal of Law in Context, 15(2), 183-197.

Mangan, D. (2021). QAnon shaman Jacob Chansley sentenced to 41 months in prison for Jan. 6 Capitol riot case. Retrieved 7 April 2022, from https://www.cnbc.com/2021/11/17/qanon-shaman-jacob-chansley-sentencing-in-trump-capitol-riot-case-.html

Massanari. # Gamergate and The Fappening: How Reddit's Algorithm, Governance, and Culture Support Toxic Technocultures. In New Media & Society. Sage Publications Sage UK: London, England, 2017.

McAleenan, K. (2019). Strategic Framework for Countering Terrorism and Targeted Violence. Department of Homeland Security

McCollister, K. E., French, M. T., & Fang, H. (2010). The cost of crime to society: new crime-specific estimates for policy and program evaluation. Drug and alcohol dependence, 108(1-2), 98–109. https://doi.org/10.1016/j.drugalcdep.2009.12.002

McDowell-Smith, A., Speckhard, A., & Yayla, A. S. (2017). Beating ISIS in the digital space: Focus testing ISIS defector counter-narrative videos with American college students. Journal for Deradicalization, (10), 50-76.

Mills Rodrigo, C. (2021). Zuckerberg to express openness to Section 230 reform | The Hill. Retrieved 5 April 2022, from https://thehill.com/policy/technology/523039-zuckerberg-to-express-openness-to-section-230-reform/

Morris, L., & Mekhennet, S. (2019). Europe has resisted taking back citizens who joined ISIS. Now, it may not have a choice.. Retrieved 30 March 2022, from https://www.washingtonpost.com/world/europe/europe-has-resisted-taking-back-citizens-who-joined-isis-now-it-may-not-have-a-choice/2019/11/14/e2137fe0-0590-11ea-9118-25d6bd37dfb1_story.html.

Moskalenko, S., & McCauley, C. (2021). QAnon. Perspectives on Terrorism, 15(2), 142-146.

Moskalenko, S., González, J. F. G., Kates, N., & Morton, J. (2022). Incel Ideology, Radicalization and Mental Health: A Survey Study. The Journal of Intelligence, Conflict, and Warfare, 4(3), 1-29.

Nihal Krishan, N. (2021). FCC nominee's record is at odds with Biden censorship goals. Retrieved 7 April 2022, from https://www.washingtonexaminer.com/policy/fcc-nominees-record-is-at-odds-with-biden-censorship-goals

O'Hara, K., & Stevens, D. (2015). Echo chambers and online radicalism: Assessing the Internet's complicity in violent extremism. Policy & Internet, 7(4), 401-422.

Office of the Director of National Intelligence. (2021) (U) Domestic Violent Extremism Poses Heightened Threat in 2021.

O'Malley, R. L., Holt, K., & Holt, T. J. (2020). An exploration of the involuntary celibate (incel) subculture online. Journal of interpersonal violence, 0886260520959625.

Papadamou, K., Zannettou, S., Blackburn, J., De Cristofaro, E., Stringhini, G., & Sirivianos, M. (2020). Understanding the incel community on youtube. arXiv preprint arXiv:2001.08293.

Percich, A. (2021). Supreme Gentlemen or Radicalized Killers: Analyzing the Radicalization Paths of Involuntary Celibate Killers and the Role of the Online Incel Forums (Doctoral dissertation, Georgetown University).

Popken, B. (2021). As algorithms take over, YouTube's recommendations highlight a human problem. N.B.C. News. Retrieved 5 December 2021, from https://www.nbcnews.com/tech/social-media/algorithms-take-over-youtube-s-recommendations-highlight-human-problem-n867596.

Raitanen, J., & Oksanen, A. (2019). Deep interest in school shootings and online radicalization. Journal of threat assessment and management, 6(3-4), 159.

Reardon, M. (2021). Section 230: How it shields Facebook and why Congress wants changes. Retrieved 1 April 2022, from https://www.cnet.com/news/politics/section-230-how-it-shields-facebook-and-why-congress-wants-changes/

Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A., & Meira Jr, W. (2020, January). Auditing radicalization pathways on YouTube. In Proceedings of the 2020 conference on fairness, accountability, and transparency (pp. 131-141).

Russonello, G. (2021). QAnon now as popular in US as some major religions, poll suggests. New York Times, 15.

Schaeffer, K (2020) "A Look at the Americans Who Believe there is Some Truth to the Conspiracy Theory that COVID-19 Was Planned," Fact Tank, Pew Research Centre, Retrieved 4 April 2022 https://www.pewresearch.org/fact-tank/2020/07/24/a-look-at-the-americanswho-believe-there-is-some-truth-to-the-conspiracy-theory-that-covid-19-was-planned/

Sganga, N. (2022). New Secret Service report details growing incel terrorism threat. Retrieved 7 April 2022, from https://www.cbsnews.com/news/incel-threat-secret-service-report/

Smith, L. (2021). In the early 1980s, white supremacist groups were early adopters (and masters) of the internet. Medium. Retrieved 5 December 2021, from https://timeline.com/white-supremacist-early-internet-5e91676eb847.

Spiccia, P. (2013). The Best Things in Life Are Not Free: Why Immunity Under Section 230 of the Communications Decency Act Should Be Earned and Not Freely Given. Val. UL Rev., 48, 369.

Stevens, and Neumann. 2012. Countering Online Radicalisation: A Strategy for Action. London: International Centre for the Study of Radicalisation. http://icsr.info/wp-content/uploads/2012/10/1236768491ICSROnlineRadicalisationReport.pdf.

Suiter, J., & Carolan, L. (2019). Radicalisation and the amplification of extremism online. Inside Politics [Podcast]. Retrieved 6 December 2021, from https://soundcloud.com/irishtimes-politics/radicalisation-and-the-amplification-of-extremism-online.

Turner, L. (2021). Sir David Amess death: Jo Cox's husband had 'physical reaction' to killing. BBC News. Retrieved 30 March 2022, from https://www.bbc.com/news/uk-58951887.

Tworek, H. (2019). Social media platforms and the upside of ignorance. CIGI, September, 9.

UNICRI :: United Nations Interregional Crime and Justice Research Institute. Unicri.it. (2021). Retrieved 6 December 2021, from http://www.unicri.it/News/-Countering-Terrorism-Online-with-Artificial-Intelligence.

United Nations. (2012). The Use of the Internet for Terrorist Purposes [United Nations Counter-Terrorism Implementation Task Force].

van der Vegt, I., Gill, P., Macdonald, S., & Kleinberg, B. (2019). Shedding light on terrorist and extremist content removal. Global Research Network on Terrorism and Technology.

Viscusi WK, Aldy J.EJ.E. The value of a statistical life: a critical review of market estimates throughout the world. J Risk Uncertain. 2003;27:5–76

Von Behr, I. (2013). Radicalisation in the digital era: The use of the internet in 15 cases of terrorism and extremism.

Vynck, G., Zakrzewski, C., Dwoskin, E., & Lerman, R. (2021). Big tech CEOs face lawmakers in House hearing on social media’s role in extremism, misinformation. Washington Post. Available online at: https://www. washingtonpost. com/technology/2021/03/25/facebook-google-twitter-house-hearing-live-updates/Mills Rodrigo, C. (2021). Zuckerberg to express openness to Section 230 reform | The Hill. Retrieved 5 April 2022, from https://thehill.com/policy/technology/523039-zuckerberg-to-express-openness-to-section-230-reform/

Ware, J. (2021). Beta Uprising. Counter Terrorist Trends and Analyses, 13(2), 10-15.

Weaver, M., & Morris, S. (2021). Plymouth gunman: a hate-filled misogynist and ‘Incel’. The Guardian, August, 13.

Weimann, G., & Von Knop, K. (2008). Applying the Notion of Noise to Countering Online Terrorism. Studies In Conflict & Terrorism, 31(10), 889. https://doi.org/10.1080/10576100802342601

Weimann, G. (2012). Lone wolves in cyberspace. Journal of Terrorism Research.

Wiktorowicz, Q. (2005). A genealogy of radical Islam: Studies in Conflict & terrorism.

Winter, C., Neumann, P., Meleagrou-Hitchens, A., Ranstorp, M., Vidino, L., & Fürst, J. (2020). Online extremism: research trends in internet activism, radicalization, and counter-strategies. International Journal of Conflict and Violence (I.J.C.V.), 14, 1-20

Witt, T. (2020). 'If i cannot have it, i will do everything i can to destroy it.'the canonization of Elliot Rodger:'Incel'masculinities, secular sainthood, and justifications of ideological violence. Social Identities, 26(5), 675-689.

Wolfowicz, M. (2015). A social learning theory examination of the complicity of personalization algorithms in the creation of echo chambers of online radicalization to violent extremism.

Wood, M. A. (2017). Antisocial media and algorithmic deviancy amplification: Analysing the id of Facebook's technological unconscious. Theoretical Criminology, 21(2), 168-185

Yokotani, K., & Takano, M. (2022). Effects of suspensions on offences and damage of suspended offenders and their peers on an online chat platform. Telematics and Informatics, 101776.

Zadrozny, B. (2021). 'Carol's Journey': What Facebook knew about how it radicalized users. N.B.C. News. Retrieved 5 December 2021, from https://www.nbcnews.com/tech/tech-news/facebook-knew-radicalized-users-rcna3581

Zorthian, J. (2021). Washington Wants to Regulate Facebook's Algorithm. That Might Be Unconstitutional. Retrieved 7 April 2022, from https://time.com/6106643/facebook-algorithm-regulation-legal-challenge/