Confronting White Supremacist Activity Online: Examining Internet Speech in the Post-Christchurch Era

by

Henry T. Honzel

A thesis submitted in partial fulfillment of the requirements
for graduation with Honors in Politics.

Whitman College
2020

Certificate of Approval

This is to certify that the accompanying thesis by Henry T. Honzel has been accepted in partial fulfillment of the requirements for graduation with Honors in Politics.

_____
Jack Jackson

Whitman College
May 13, 2020

Table of Contents

Acknowledgements

I'd like to thank my family for the constant support they've given me throughout not just this project but the four years I've had at Whitman – it has truly meant the world and I could not have done any of this without you guys.

I'd also like to extend thanks to the incredible professors who have guided me through my time at Whitman:

Professor Jackson, who constantly challenged my thinking and helped me discover my interest in law.

Professor Apostolidis, who served as my introduction into the Whitman Politics program and into the broader world of politics as a whole.

Professor M. Acuff, who urged me to take my first politics class as an incoming First Year, something I will forever be grateful for on so many levels.

Finally, I'd like to thank my friends: Grant, Perth, Trent, Sarah, Kate, Georgia, Darby, Declan, Ben, David, Walter, the list goes on. Your patience and your support have been incredible and have helped me through some of the most challenging moments of my life. Thank you, truly.

Introduction

On March 13, 2019, a livestream video surfaced on Facebook broadcasting from the city of Christchurch, New Zealand. Captured from a small helmet-mounted camera, the video depicts a white male in his twenties driving toward a local mosque covered in military-style gear. As he arrives, he exits the car and draws a weapon covered in scrawled inscriptions bearing reference to a wide variety of fringe white supremacist internet posts and slang. Before entering the mosque, he speaks directly to the audience that has digitally amassed to watch the livestream by quoting a phrase associated with the wildly popular YouTuber Felix Kjellberg, a figure who has generated controversy for a number of high profile bigoted remarks and risen as a form of idol in online white supremacist groups: "Remember, lads, subscribe to PewDiePie."[1] With this less than subtle nod to the deep online networks that radicalized him, he advances toward the mosque and begins to fire.

The final video ran a total of 17 minutes and would capture in horrific detail the first segment of a mass shooting that would leave 49 people dead. Following its initial upload to Facebook, the livestream would go on to spread like wildfire across the digital landscape with countless re-uploads appearing across major online platforms including Twitter, Reddit, and YouTube.[2] Released prior to the attack, a 74 page manifesto written by the perpetrator and uploaded to 8chan—an online forum notorious for its ties to white

1 Kevin Roose, "A Mass Murder of, and for, the Internet," *The New York Times,* March 15, 2019, https://www.nytimes.com/2019/03/15/technology/facebook-youtube-christchurch-shooting.html
2 Ibid.

supremacists groups—also began to circulate on a variety of social media sites and forums.

Within hours of the video's upload, Facebook and other prominent online platforms had begun the process of identifying, isolating, and deleting any copies of the video, manifesto, and fake accounts claiming to be the perpetrator that had cropped up following the attack.[3] Although most platforms managed to remove the vast majority of video copies in less than a day, the immense scale and speed of transmission facilitated by the internet was readily apparent. On Facebook alone, 1.5 *million* copies of the video were removed from the site within the first 24 hours,[4] with hundreds more taking root in the darker, less regulated portions of the internet.

As news of the attack spread on a global stage and media outlets began to dig into the shooter's manifesto, it became increasingly clear that the perpetrator sought for this attack to serve as a continuation in a long lineage of prior white supremacist attacks. While the usage of Facebook's livestream technology to directly broadcast the attack itself presented an appalling new twist, the Christchurch attack was certainly not the first to turn to online forums and social media sites as a method to publicize white supremacist violence and attempt to persuade others to commit similar actions. Originating with the 2011 mass shooting in Norway that left 77 people dead, a network of interconnected attacks has emerged across the globe, with perpetrators in the United States, Germany,

3 Kate Klonick, "Inside the Team at Facebook That Dealt with the Christchurch Shooting," *The New Yorker,* April 25, 2019, https://www.newyorker.com/news/news-desk/inside-the-team-at-facebook-that-dealt-with-the-christchurch-shooting

4 Ibid.

Sweden, France, New Zealand, and more all citing one another as influences for their actions in a litany of online posts, manifestos, and videos.[5]

With this deadly trend showing little sign of abating, this project seeks to answer two central questions: Is white supremacist speech in online spaces something that needs to be confronted and potentially censored, and if so, how do we as a society go about doing so in a manner that is effective, just, and transparent? In my first section, I argue that given the immense amplifying power of the internet as a communication tool coupled with the frequency and severity of the attacks committed, white supremacist speech online *does* indeed constitute a rather dire threat that must be addressed. In acknowledging that such speech does pose a genuine threat, in my second section I then examine the benefits and drawbacks to two approaches towards regulation: one that relies on a law-centric approach rooted in public power, and another that relies on a corporate-centric approach rooted in private power. To conclude, I finally argue that the best method through which to regulate white supremacist speech online is a hybrid approach between the two aforementioned methods, one that is capable of adjusting to scenarios that don't align directly with prior legal precedent yet additionally avoids engaging in excessive and arbitrary censorship.[6]

5 Weiyi Cai and Simone Landon, "Attacks by White Extremists Are Growing. So Are Their Connections." *The New York Times,* April 3, 2019, https://www.nytimes.com/interactive/2019/04/03/world/white-extremist-terrorism-christchurch.html

6 It should be noted that although this paper deals with case studies such as the Christchurch attack that occurred internationally, this project seeks to treat these instances and any regulatory approaches within the frame of U.S. Jurisprudence

Part I - Is White Supremacist Speech Online a Real Threat?

A substantial portion of assessing the threat posed by white supremacist speech online is understanding the way in which their online community is organized and the methods they employ to communicate their ideology both internally and externally. While there is notably no single definitive organization that binds together online white supremacist communities, a wide variety of far-right groups have come to categorize themselves as members of the "alt-right." First emerging in 2008 as a creation of the infamous white nationalist Richard Spencer, the "alt-right" began as a way to define the loose network of "far-right ideals centered on 'white identity' and the preservation of 'Western civilization."[7] By 2010, Spencer began refining the foundational principles of the movement with the launch of the "Alternative Right" blog that focused primarily on a refutation of mainstream conservatism and a heavy emphasis on the recruitment and radicalization of primarily younger internet-savvy individuals.[8] Although the blog itself would eventually fall into obscurity following Spencer's abandonment of the project in 2013, momentum behind the alt-right movement was only beginning to build.

While numerous figures including Jared Taylor (editor of the pseudo-science *American Renaissance* journal), Greg Johnson (from the publishing house Counter-Currents), Mike "Enoch" Peinovich (founder of *The Right Stuff* blog), and Richard Spencer himself have at some point proclaimed themselves to be "leaders" of the alt-

7 "Alt-Right," Southern Poverty Law Center, accessed February 18, 2020, https://www.splcenter.org/fighting-hate/extremist-files/ideology/alt-right
8 Ibid.

right, the movement itself is quite diffuse and is primarily composed of "anonymous youths who were exposed to the movement's ideas through online message boards like 4chan and 8chan's /pol/ and Internet platforms like Reddit and Twitter."[9] As evidenced by the composition of their base, the internet is a crucial tool for the alt-right. As Richard Spencer recognized in his original blog, online spaces are well tailored to the process of recruitment and radicalization, especially for a movement that relies on ideology that has largely been pushed out of traditional public spaces. Online forums and social media sites provide a robust blend of factors that make communication fast, effective, and—perhaps most pivotally—challenging to regulate. First, the vast constellation of websites and the dispersed global nature of their visitors makes it nearly impossible to clearly identify a specific online audience for any given post, picture, or video.[10] Second, the speed of communication across forums, messaging applications, and social media sites means material can be transmitted amongst hundreds of thousands of people nearly instantly.[11] Third, the ability to mask one's real identity through usernames and online handles gives individuals a sense of anonymity and protection.[12] Finally, the low cost of access and the ubiquity of smart devices allows nearly anyone to access the internet at practically any point.[13]

On the one hand, these factors make the internet a quite unique platform, offering a place to voice one's opinions quickly to a wide audience without fear of judgement

9 Ibid.

10 Scott Hammack, "The Internet Loophole: Why Threatening Speech On-line Requires a Modification of the Courts' Approach to True Threats and Incitement," *Columbia Journal of Law and Social Problems* 36, no. 1 (Fall 2002): 12.

11 Hammack, "The Internet Loophole," 13.

12 Ibid.

13 Hammack, "The Internet Loophole," 14.

based on one's identity.[14] Unfortunately, these same factors that make the internet such a haven for free speech also greatly amplify the threat posed by speech that seeks to incite violence or convey a threat.

To better conceptualize this process of amplification, it is useful to turn to an example posed by John Stuart Mill in his piece *On Liberty.* Published in 1859, *On Liberty* existed far from the internet-laden context of today, yet Mill struggled with an issue that remains in contemporary society today: presuming that we have the power to regulate speech, how do we identify and restrict speech that may seek to cause harm *without* infringing on the ability for an individual to express a critical opinion? In seeking to determine what speech should be deemed acceptable and what should not, Mill explains that "An opinion that corn-dealers are starvers of the poor, or that private property is robbery, ought to be unmolested when simply circulated through the press, but may justly occur punishment when delivered orally to an excited mob assembled before the house of the corn-dealer, or when handed about among the same mob in the form of a placard."[15]

In this example, the *space*, *method of delivery,* and *audience* for the speech are of the utmost importance. Although the opinion about a figure such as the corn-dealer might be the same in both cases, the two instances differ substantially in the message they deliver and the effect they ultimately wish to produce. When these aforementioned elements are clearly observable, the likelihood of discerning the magnitude of the presented threat is much easier. In the case of the example posed by Mill it is easy to see how a passionate speech to an excited mob decrying the corn-dealer in front of their

14 Ibid.

15 John Stuart Mill, "Of Individuality, as One of the Elements of Well-Being," in *On Liberty and Other Writings,* ed. Stefan Collini (Cambridge: University Press, 1989), 57.

home would likely result in an outcome of harmful or violent action. Determining what speech is merely an opinion and what speech seeks to incite others towards destructive action thus becomes a sort of puzzle in which contextual factors must be sorted through to ultimately deduce what the intent behind the speech might be.

With the internet, however, these contextual factors become far murkier. The ability to disseminate media to a global audience instantly and anonymously makes it incredibly difficult to gauge the severity of a threat and even more difficult to interpret the likelihood of it being carried out. Without a clear sense of the vital contextual factors, the line between a fair but critical opinion, a pointed satire opinion, and a malicious threatening opinion becomes incredibly difficult to judge. This threat amplification in itself provides reason for alarm and indicates the radically transformative nature of the internet on speech, yet the potential for abuse does not indicate a need to engage in regulation and censorship on its own, and many organizations utilize the power of the internet in a positive way.

Regulation and censorship *does* become a necessity, however, once one begins to examine the emergent pattern of interlinked white supremacist attacks that emerge directly as a product of radicalizing online communities. To understand how these attacks emerge, one must first trace how white supremacist communities online are capable of recruiting and radicalizing members to such a degree that they feel not only motivated but *obligated* to commit acts of mass violence. Forgoing more formal recruitment techniques, one of the most pivotal methods of communication utilized by the alt-right to spread their ideology are memes, a type of media that emerged as a product of online forums and social media sites. Best described as "user-generated media that share

recognizable characteristics in content or form through which their creators seek to guide viewers' interactions and interpretations,"[16] memes allow caustic white supremacist ideology to be disseminated in a manner that on the surface appears satirical, ironic, and "humorous." The co-opting of seemingly innocuous figures such as a cartoon frog named Pepe[17] allow white supremacist groups to spread intensely bigoted views behind the cover of childish cartoons.

Weaponizing satire, irony, and "humor" inherent to the meme format has become perhaps the alt-right's most critical tactic in pushing what were once "backstage" ideologies discussed only in tightly enclosed white supremacist circles onto the "frontstage" of mainstream discourse."[18] As stated in a leaked portion of a style guide for the notorious white supremacist website *The Daily Stormer:* "Packing our message inside of existing cultural memes and humor can be viewed as a delivery method. Something like adding cherry flavor to children's medicine."[19] By co-opting a mainstream meme format like Pepe the Frog that frequently appears on popular websites and imbuing them with an extreme and outrageous far-right message, alt-right groups not only push their ideologies into the mainstream in a relatively covert manner, but further add a degree of deniability following the recognition of the meme's actual toxic content.[20] Having

16 Viveca S. Greene, "Deplorable' Satire: Alt-Right Memes, White Genocide Tweets, and Redpilling Normies," *Studies in American Humor* 5, no. 1 (2019): 38. Project MUSE.

17 "Pepe the Frog," General Hate Symbols, Anti-Defamation League, accessed February 29, 2020, https://www.adl.org/education/references/hate-symbols/pepe-the-frog

18 Greene, "Deplorable' Satire," 36.

19 Greene, "Deplorable' Satire," 54.

20 Whether or not one considers this widespread dissemination of "hate speech" problematic is the subject of a long standing debate. This project considers hate speech in largely the same vein as the theorist Catherine Mackinnon who emphasizes the way in which the protection of offensive and bigoted ideologies reinforces oppressor/oppressed dynamics in society and therefore believes it should be regulated in favor of principles of greater equality *Only Words* (Cambridge: Harvard University Press, 1993,) 75. This view is not universally held however, and as theorist Henry Louis Gates Jr. argues that we have come to overvalue the weight of words and therefore come to sacrifice liberty without actually securing equality merely

emerged in the informal realm of internet forums, memes are often "dismissed as 'innocent, comical, and even cute."[21] They are frequently viewed as jokes that carry no real significance, something that provides a layer of defense for those who utilize them to express hateful ideologies. Upon receiving accusations that a meme they created is offensive, hateful, or threatening, the creator immediately claims that the work was only done in jest and is far too ridiculous to be considered a part of serious discourse. This built-in deniability is by design, actively shielding the meme's creator from potential repercussions.

Further, the use of popular meme formats to spread white supremacist ideology creates a sense of community among individuals who engage in internet spaces with such content frequently enough to become "in" on the joke.[22] While most people would only see a crudely drawn cartoon frog upon seeing a meme of Pepe, those who frequent white supremacist groups immediately recognize the symbol as a form of bigoted dog whistle. This in turn draws in individuals seeking validation and a sense of belonging,[23] pushing them to begin creating their own renditions of such memes in hopes of receiving praise and being accepted into these white supremacist communities. As individuals bond over their collective understanding of this predatory content, they increasingly view themselves as separate from those who are offended by such memes.

As these online white supremacist communities crystalize and isolate, the content they produce begins to become increasingly radical as they seek to outdo one another.

through the elimination of offensive rhetoric and ideologies. For more on this divide it is useful to explore Mackinnon's book *Only Words* and Gates' essay "Let Them Talk."

21 Greene, "Deplorable' Satire," 42.

22 Greene, "Deplorable' Satire," 47.

23 Hammack, "The Internet Loophole," 12.

This ceaseless cycle of radicalization eventually pushes individuals into acting upon their ideologies as a means of standing out within these communities. Oftentimes, as was the case in the Christchurch attack, these individuals publish extensive manifestos as a method of constructing themselves as self-fashioned martyrs. These manifestos seek to solidify the perpetrator's status as an "icon" amongst the white supremacist communities that radicalized them and call upon others to join the ranks of those who have taken action. Manifestos thus often serve as the network that links white extremist actors across the globe, with a continually growing chain emerging with each subsequent attack as each subsequent attacker reveres and idolizes those that came before them. The result is the emergence of a disturbing lineage of attacks: Five different extremists (including the perpetrator of the Christchurch attack) have directly cited the 2011 bombing and mass shooting committed by a white extremist in Norway as an influence for their own attacks.[24] Likewise, a recent analysis found that "at least a third of white extremist killers since 2011 were inspired by others who perpetrated similar attacks, professed a reverence for them or showed an interest in their tactics."[25]

The emergence of this trend demonstrates a clear message: the potential for the internet to serve as an amplifying factor for threatening and inciting speech has become a deadly reality in the case of white supremacist communities online. Utilizing the ability to widely disseminate toxic ideology in an anonymous fashion, white supremacist groups and the alt-right movement have managed to create online communities that have repeatedly radicalized individuals to the point of committing horrific attacks. Capitalizing

24 Cai and Landon, "Attacks by White Extremists Are Growing."
25 Ibid.

on the difficulty to regulate such spaces, these communities have further manifested a cycle of violence that is revitalized and even heightened with each attack and shows little sign of abating. In returning then to the question "is white supremacist speech online a real threat?" I would argue yes it is, and a quite substantial one at that.

With this recognition in mind, the question we are thus presented with is whether we should pursue regulatory action against white supremacist speech online, and if so what form such regulation may take. This question is anything but simple, and its answer may resemble more of a spectrum than a singular definitive point. On perhaps a more extreme end, it could be argued that the majority of the content produced by these white supremacist group (excluding the most extreme instances such as the actual video documentation of the Christchurch attack) is actually expressive discourse that is *pivotal* to the function and health of a democratic society. Articulated in the majority opinion of *Terminiello v. Chicago,*[26] Justice William O. Douglas stated that a crucial function of free speech is actually to "invite dispute."[27] Justice Douglas further argues free speech and discourse "may indeed best serve its high purpose when it induces a condition of unrest, creates dissatisfaction with conditions as they are, or even stirs people to anger."[28] Under this understanding one could argue that memes, forum posts, and even certain components of manifestos are thus actually productive components of civil discourse that—despite posing the threat of potential violence and unrest—should largely be protected exactly *for* this reason, outside of the most extreme cases.

26 The case concerned the conviction of Father Arthur Terminiello, a Catholic priest who was arrested for causing a "breach of the peace" during a speech he gave in a Chicago auditorium. The speech involved the substantial criticism of various racial and political groups and resulted in a large gathering of protestors outside of the auditorium that police were unable to control.

27 Terminiello v. Chicago, 337 U.S. 1, 4, (1949).

28 Terminiello v. Chicago, 337 U.S. 1, 4, (1949).

Another approach to solving this question would be to argue against the notion that the internet has changed the nature of inciting content in any substantial manner. Central to this approach is the thought that although it may be easier to mock someone through the internet or spread hateful ideologies, the actual standards for determining threats and incitement have not changed. Since such standards remain static, it could be argued that "moving the line" of acceptable free speech alongside the adoption of new technology opens a dangerous path towards continually expanding censorship.[29] This approach essentially refutes the argument made earlier in this paper that communication forms unique to internet spaces such as memes or online forum posts alter the nature of what might constitute inciting speech and states that such forms do not pose an inherently greater risk for the enactment of violent action and therefore do not deserve special consideration.[30] Under both this approach and the one presented in *Terminiello* one could argue that regulatory action against white supremacist speech online would be either unnecessary or perhaps even counter-productive to discourse as a whole. These approaches subscribe rather heavily to the concept of a "marketplace of ideas" in which the only remedy to the expression of ideologies that one considers false, hateful, problematic, etc. is the employment of speech designed to counter such ideologies.[31]

While these methods present viable options, Judith Butler presents a compelling argument in her piece "Limits of Free Speech?" that pushes for the necessity of taking regulatory action in the face of newly emergent online threats. Responding to an incident that took place on the University of Wisconsin-Milwaukee campus in which a prominent

29 John K. Wilson, "In Opposition to Butler's 'Limits on Free Speech," Academe Blog, December 12, 2017, https://academeblog.org/2017/12/12/in-opposition-to-butlers-limits-on-free-speech/
30 Ibid.
31 Catherine Mackinnon, *Only Words* (Cambridge: Harvard University Press, 1993) 75.

alt-right figure utilized cameras to nonconsensually broadcast a trans student during a speech and actively shamed them while encouraging others to follow suit, Butler argues that the spread of new technologies "produce new possibilities for incitement, harassment, and the commission of illegal activities."[32] Although the fact pattern present in this specific incident does not necessarily align with the broader context of white supremacist speech online, the notion that the introduction of technology *can* change our understanding of incitement and "force us to reconsider the meaning of expressive freedom" is critical.[33] Butler further argues that, unlike the two prior approaches, we are not necessarily bound to ultimately deferring to the First Amendment over all other values.[34] This is not to say that the First Amendment must be discarded, and Butler makes it clear that she herself is a free speech advocate, but rather that we must remain vigilant about how new forms of communication may disrupt our prior demarcations of acceptable and unacceptable speech and actively pursue regulation in the moments when such new forms present serious threats to values such as the dignity and safety of others.[35] I would argue that the continual expansion of white supremacist violence has suggested that such a realization and subsequent pursuit of action is long overdue. Yet, in taking action, we must remain vigilant of the tremendous power that can come with censorship and regulation, and seek to do so in a way that does not lose sight of the values of transparency and justice. The next section of this project thus seeks to analyze

32 Judith Butler, "Limits on Free Speech?," Academe Blog, December 7, 2017, https://academeblog.org/2017/12/07/free-expression-or-harassment/

33 Ibid.

34 Ibid.

35 Ibid.

two methods through which to approach this problem and identifies the strengths and weaknesses present within both.

Part II - Regulating White Supremacist Speech Online

Section A - Engaging in a Law-centric Approach

Traditionally, the process of identifying and regulating speech (within the United States) that is classified as inciting or threatening has been left to the discretion of the legal system. Although the First Amendment of the U.S. constitution states that "Congress shall make no law...abridging the freedom of speech,"[36] courts have long avoided adhering to such an absolutist conception in their interpretations of cases. As Butler mentions in her article, one of the great legal challenges in U.S. jurisprudence is seeking to strike a balance between upholding the principle of free speech without necessarily sacrificing other constitutional values such as equal protection.[37] The result of this conceptual struggle is the emergence of a delicate legal balancing act that attempts to hold all of these principles in line with one another. Perhaps more notably, however, is that this balancing act has shifted over time as different eras of the courts produce different interpretations and seek to tip the balance in different directions. To understand the strengths and weaknesses of a law-centric approach to white supremacist speech

36 U.S. Const. amend. I
37 Butler, "Limits on Free Speech?"

online, it is thus important to first walk through a legal history of incitement cases to understand how precedent has evolved and how it stands today.

On the most basic level, *to incite* is defined as "to move to action : stir up : spur on : urge on"[38] The primary foundations for U.S. jurisprudence related to incitement came in 1919 with the Supreme Court case of *Schenck v. United States.* The defendants in the case were convicted under the Espionage act following their arrest for distributing leaflets that urged individuals to resist the ongoing World War I draft.[39] The defendants appealed their convictions under the claim that the Espionage act violated their First Amendment rights to both freedom of speech and freedom of press.[40] The Supreme Court voted to uphold the defendants convictions, with the majority opinion by Justice Oliver Wendell Holmes establishing the basis through which future cases of incitement to be interpreted. Initially, Justice Holmes clarifies that there are some cases of speech that should never receive constitutional protection, such as that of an individual shouting fire in a crowded theatre with the intent of causing panic.[41] Yet not all cases are necessarily so clear, and therefore in dealing with potential instances of incitement Justice Holmes called upon the court to consider "whether the words used are used in such circumstances and are of such a nature as to *create a clear and present danger* [emphasis added] that they will bring about the substantive evils that Congress has a right to prevent"[42] as a means in which to "test" speech to qualify it as incitement.

38 "incite" *Merriam-Webster.com.* 2020. https://www.merriam-webster.com (13 February 2020).
39 Schenck v. United States, 249 U.S. 47, 49, (1919).
40 Schenck v. United States, 249 U.S. 47, 50, (1919).
41 Schenck v. United States, 249 U.S. 47, 52, (1919).
42 Ibid.

In that same year, the utilization of the clear and present danger test (henceforth referred to as the C&PD test) would be solidified by the Supreme Court case of *Debs v. United States (1919).* Similar to *Schenck, Debs* involved the advocacy by an individual (in this case prominent socialist and presidential candidate Eugene V. Debs) against American involvement in World War I. Debs additionally offered praise to Kate Richards O'Hare, an individual recently convicted for obstructing the enlistment service.[43] In his opinion, Justice Holmes states specifically that Debs could not be found guilty merely for "advocacy of any of his opinions *unless the words used had as their natural tendency and reasonably probable effect to obstruct the recruiting service, &c., and unless the defendant had the specific intent to do so in his mind.* [emphasis added],"[44] effectively restating the logic of the C&PD test. Believing that such criteria *was* met, the court upheld Debs' conviction on the grounds that his speech *did* in fact in their interpretation provide a genuine threat to the recruitment process.

As was the case with the example of the corn-dealer posed by Mill, the court's decision in both *Schenck* and *Debs* relied heavily on the presence of contextual factors such as the method of speech presentation and the audience the speech was delivered to. Notably, both cases were also substantially shaped by the greater context of the U.S.'s involvement in World War I. Perceiving a substantial threat to national security, the courts sought to act in deference to the preservation of U.S. military institutions over the principle of pure free speech. This deliberate act of deference illustrated some of the major concerns that arose alongside the adoption of the C&PD test, most notably that it

43 Debs v. United States, 249 U.S. 211, 39, 212-213, (1919).
44 Debs v. United States, 249 U.S. 211, 39, 216, (1919).

offered a great potential for abuse and excessive censorship. The test established a rather low level of scrutiny through which a speaker could be convicted and particularly in periods of crisis or conflict (such as that of *Schenck* and *Debs*) could be easily mobilized to suppress speech critical of government institutions.

Even prior to the formal adoption of the C&PD test in *Schenck* and *Debs,* fears regarding the potential for instances of excessive censorship under the guise of incitement regulation existed. Perhaps the most notable critique and the basis through which future criticism of the C&PD test would emerge was the opinion of District Judge Learned Hand in the case of *Masses Publishing v. Patten (S.D.N.Y.) (1917).* Despite being situated in the same context of American involvement in World War I, Hand offered an adamant defense of free speech by advocating

against the conviction of the magazine *Masses* for their publication of cartoons depicting strong anti-war sentiment. Hand recognized that situations of conflict and unrest such as war frequently increased the desire to tip the judicial balance away from principles of free speech, yet warned that such an urge must be resisted as free speech served as a necessary "safeguard of free government."[45] Seeking to protect what he saw as a fundamental cornerstone of American government, Hand stated that the court must distinguish between "keys of persuasion" and "triggers of action."[46] He felt that while the cartoon certainly demonstrated support for those who resist the draft, a demonstration of support is *not* equivalent to actual advocacy for resistance.[47] Under this important

45 Masses Publ'g v. Patten, 244 F. 535, (S.D.N.Y. 1917).

46 John R. Vile, "*Masses Publishing Co. v. Patten* (S.D.N.Y) (1917)," *The First Amendment Encyclopedia,* accessed February 16, 2020, https://www.mtsu.edu/first-amendment/article/502/masses-publishing-co-v-patten-s-d-n-y

47 Vile, "*Masses Publishing Co. V. Patten,"*

separation, Hand felt that the cartoons published in *Masses* should therefore be considered a protected form of speech as they only sought to persuade others of an opinion, *not* push them towards direct action.[48]

Although Judge Hand's decision would ultimately be overturned, the concerns he brought forth in *Masses* would come to dramatically influence the later decisions of Justice Holmes, the creator of the C&PD test. Recognizing the potential for the test to be mobilized in a manner that produced excessive censorship, Holmes' later case opinions reflected an increased desire to offer greater protection to speakers. In the cases of *Abrams v. United States (1919) and Gitlow v. New York (1925)* Justice Holmes would dissent against the court's majority opinion, each time asserting that the defendant's actions did not truly constitute a clear and present danger and instead were equivalent merely to protected critical opinions. Additionally, these dissents served as the beginnings of a decades long process of refinement and alteration of the C&PD test. In *Abrams,* Holmes added a temporal component, asserting that not only must speech indicate a clear and present danger, but said danger must be of an "immediate evil."[49] In *Gitlow,* Holmes quite radically states that "every idea is an incitement," however further asserts that this does not mean that all potentially dangerous or radical ideas must be suppressed, particularly if they do not have the actual genuine support to be carried out.[50]

The most direct critique against the C&PD test, however, was not actually directly penned by Holmes, but rather came from Justice Louis Brandeis in his concurring opinion in *Whitney v. California (1927)* that Holmes joined. *Whitney* involved an

48 Vile, "*Masses Publishing Co. V. Patten,"*
49 Abrams v. United States, 250 U.S. 616, 627-628, (1919).
50 Gitlow v. New York, 268 U.S. 652, 673, (1925).

individual who was part of the Communist Labor Party of California and was prosecuted under the California Criminal Syndicalism Act for her alleged involvement in attempts to affect "economic and political change through the unlawful use of violence."[51] The defendant attempted to challenge the Syndicalism act on the basis that it violated her First Amendment protections, which the court unanimously decided was not true. In his opinion, Brandeis concurs with the majority opinion's ruling on the basis that free speech is not an absolute right and that there do exist instances in which governmental power may intervene when speech seeks to "incite to crime, disturb the public peace, or endanger the foundations of organized government and threaten its overthrow by unlawful means."[52] Yet more profoundly, Brandeis urges that given that the right to free speech is interpreted as being non-absolute, substantial caution must be used going forward. He quite starkly asserts that it is abundantly evident that the court has not been capable at arriving at a true fixed standard to assess the severity or imminence of a given threat, and that attempting to create such measures must be done with a deliberate deference to *permitting* most speech rather than censoring it. In the end, Brandeis argues that the remedy to preventing further instances of violence against the State is "more speech, not enforced silence."[53]

By 1969, the minority opinion of Brandeis and Holmes would become the majority opinion in the pivotal case of *Brandenburg v. Ohio.* The case involved a collection of videos taken by a Cincinnati news crew of a KKK rally at a farm in the rural county of Hamilton. The videos depict various acts of the rally including a cross burning,

51 "Whitney v. California," Facts of the Case, Oyez, accessed February 16, 2020, https://www.oyez.org/cases/1900-1940/274us357

52 Whitney v. Cal., 274 U.S. 357, 371, (1927).

53 Whitney v. Cal., 274 U.S. 357, 377, (1927).

a speech by the group's leader claiming the need for "revengeance" against government members if the alleged "suppression of white Caucasians" continues, and features images of various KKK members holding firearms (although not the leader himself). The leader of the group was then convicted under the Ohio Syndicalism Statute for "advocat[ing] . . . the duty, necessity, or propriety of crime, sabotage, violence, or unlawful methods of terrorism as a means of accomplishing industrial or political reform" as well as for "voluntarily assembl[ing] with any society, group, or assemblage of persons formed to teach or advocate the doctrines of criminal syndicalism."[54]

Similar to the defendant in *Whitney,* the KKK leader challenged the Ohio Syndicalism Statute on the grounds that it violated his First Amendment protections. Unlike the ruling in *Whitney,* however, the court agreed, and in a dramatic reversal from the unanimous decision of *Whitney* ruled that the Act *was* in fact in violation of the defendant's First Amendment rights. In the per curiam majority opinion, the court essentially stated that the Ohio Syndicalism Act was too broad in defining what types of actions that it would punish, and that only "incitement to *imminent lawless action* [emphasis added]"[55] constituted unprotected speech. This profound overturn marked a substantial shift in the court's orientation towards incitement cases and largely spelled the end for the C&PD test in favor of the far more speaker-oriented imminent lawless action (henceforth ILA) test. Under this new test, not only must offending speech seek to produce imminent lawless action, but must also do so in a manner "likely to incite or to produce such action."[56] Under such a strict level of scrutiny, the ILA provides far greater

54 Brandenburg v. Ohio, 395 U.S. 444, 445, (1969).
55 Brandenburg v. Ohio, 395 U.S. 444, 449, (1969).
56 Brandenburg v. Ohio, 395 U.S. 444, 447, (1969).

protection against censorship than that of the C&PD test, and demonstrates yet another substantial rebalancing of values undertaken on behalf of the court, this time in deference to the protection of free speech.

Notably, both the greatest strength and the greatest weakness of the law-centric approach to white supremacist speech online hinges upon the continued contemporary reliance on the precedent established in *Brandenburg* to identify inciting speech. In looking first toward the strength provided, the rigorous level of protection afforded to speakers under the ILA test is of the utmost importance. As stated by Judge Hand in his opinion in *Masses,* it is in periods of crisis and conflict that the desire to censor is *most* prevalent, yet these same periods are also the points at which free speech is most important to protect.[57] This is not to say that our ultimate deference to free speech must be absolute. As revealed by Butler's piece in the previous section, such absolutism is quite dangerous. While I don't believe we should make drastic concessions to free speech that allow for the continued flourishing of harmful white supremacist speech online, I would side with Justice Brandeis in urging substantial caution when approaching censorship.[58]

Whilst it is tempting to forgo established precedent—especially when it relies on such a high standard of scrutiny such as that of the ILA and attempts to regulate content as toxic as that found in the white supremacist community—in favor of more aggressive and immediate regulation, doing so opens the potential for increasingly excessive instances of censorship. As Brandeis stated, enforcing silence is unlikely to resolve further instances

57 Masses Publ'g v. Patten, 244 F. 535, (S.D.N.Y. 1917).
58 Whitney v. Cal., 274 U.S. 357, 377, (1927).

of violence, and recklessly allocating power to the State to engage in censorship can quickly result in every instance of crisis being viewed as a potential to suppress critical opinions.

The greatest strength of a law-centric approach thus lies in the caution developed over its extensive history. As evidenced by a walkthrough of U.S. jurisprudence related to incitement, the deep-set speaker protections afforded by the ILA test is a deliberate choice made on behalf of the courts in direct response to the underwhelming protection offered by the C&PD test. This process of balancing values of free speech with values of security and dignity cannot be entirely eschewed, and in seeking to identify the best course to approach regulating white supremacist speech online it remains vital to resist the temptation to entirely forgo speaker-based protection.

Yet while the continued reliance on *Brandenburg* precedent provides a critical strength to the law-centric approach, so too does it pose a fundamental weakness. Decided in the late 1960's, *Brandenburg* existed in a context far prior to that of the internet era. As mentioned in part one of this project, the internet has drastically altered the nature of communication and allowed media and information to be disseminated instantly, anonymously, cheaply, and to a broad, nebulous audience. This alteration has increasingly resulted in an emergence of incitement-style cases with fact patterns that don't neatly conform to the imminent lawless action standards imposed in *Brandenburg.* One such example is the case of *United States v. Turner (2013)* in which Harold Turner—a talk radio show host with a large white supremacist following—published a blog post on his website that claimed that three judges involved with a Seventh Circuit court decision claiming that the Second Amendment doesn't apply to the states "deserve

to be killed for their decision."[59] Turner then additionally posted the names and photographs of the three judges alongside "a photograph and map of their courthouse marked to show the location of "[a]nti-truck bomb barriers," and the room numbers of their chambers."[60] Turner was then convicted under a statute criminalizing the threatening of a federal judge. This conviction was subsequently upheld by the Second Circuit court following Turner's appeal, but was notably done so under a true threats framework, a category that is doctrinally distinct from incitement. By classifying Turner's action as a *threat* rather than as incitement, Turner's actions were judged on the far "less speech protective and more ambiguous"[61] legal standard for true threats. This standard asks "whether an ordinary, reasonable recipient who is familiar with the context of the [communication] would interpret it as a threat of injury."[62]

The court's choice to interpret *Turner* under a threat framework demonstrates the growing challenge to transition *Brandenburg* precedent to the internet era. As communication over the internet "magnifies the problems courts face in applying *Brandenburg* to settings in which the audience is hazily defined and the meaning of 'imminence' is likewise unclear,"[63] Courts are increasingly seeking to avoid engaging with *Brandenburg* altogether. This growing notion of obsolescence speaks to one of the primary weaknesses present in a law-centric approach: it lacks flexibility in its application. Two major elements of internet era communication are speed and adaptability, both elements that a law-centric approach largely lacks. Relying on decades-

59 "First Amendment — Freedom of Speech — Second Circuit Affirms Threats Conviction in Internet Speech Case — United States v. Turner," Harvard Law Review 127, no. 8 (June 2014): 2585. JSTOR.
60 "United States v. Turner," 2586.
61 "United States v. Turner," 2589.
62 "United States v. Turner," 2586-2587.
63 "United States v. Turner," 2592.

old precedent that is tremendously difficult to align with non-conforming fact patterns has left courts struggling to resolve cases like *Turner,* and has truncated further development of incitement doctrine. This stagnation in turn has raised questions about how effective a law-centric approach may be going forward.

A law-centric approach is thus far from a perfect method through which to regulate white supremacist speech online, however it does provide valuable components. Although the ILA test remains difficult to apply to online speech that lacks a clearly defined audience and an easily identifiable temporal component, its goal to preserve some element of speaker protection is incredibly important. This is not to say that the rhetoric or ideology expressed in white supremacist communities online should be validated or accepted, but rather that speech online should be entitled to some form of standardized review process. By drawing on concrete established precedent, a law-centric approach helps to ensure that speech cannot be censored or punished without first being evaluated in a transparent manner that includes recourse via appeal should one disagree with said evaluation. Lacking the ability to adapt quickly however, it is challenging to envision a law-centric approach as the primary method of regulation on a medium as fast-paced as that of the internet. It is due to this weakness that many have increasingly come to call for an alternative approach, one that relies on the corporations and entities that control the vast array of forums and websites utilized by white supremacist groups to take action.

Section B - Engaging in a Corporate-centric Approach

Frustrated with the inability of a law-centric approach to provide a flexible and proactive response to the recurrent instances of white supremacist violence rooted in online spaces, a desire for an alternative approach has heightened with each new attack. Seeking to directly regulate the online spaces in which white supremacists congregate, recruit, and radicalize, calls have been made urging action on behalf of the corporations and individuals behind the internet platforms themselves. Mainstream sites such as Facebook, Twitter, and Reddit have encountered mounting demands to take responsibility for the content disseminated on their platforms, while sites notorious for their substantial white supremacist presence have been challenged to clean up or face being shut down entirely.[64]

These calls for a corporate-centric approach to regulation aren't new, nor do their origins rest specifically with instances of white supremacist violence. Major platforms such as YouTube have long dealt with their platforms being utilized by radical groups as a method to promote their ideologies. In 2015, YouTube had over 40,000 videos from jihadist terrorist groups,[65] a problem that only continued to grow alongside the rise of

64 Kevin Rose, "'Shut the Site Down,' Says the Creator of 8chan, a Megaphone for Gunmen," *The New York Times,* August 4, 2019, https://www.nytimes.com/2019/08/04/technology/8chan-shooting-manifesto.html

65 Jonathan Taplin, "How to Force 8Chan, Reddit, and Others to Clean Up," *The New York Times,* August 7, 2019, https://www.nytimes.com/2019/08/07/opinion/8chan-reddit-youtube-el-paso.html.

ISIS.[66] These videos frequently showed instances of extreme violence such as the infamous beheadings of two American journalists in response to continued American bombing campaigns against ISIS targets in Iraq.[67]

Although these specific uploads were quickly removed, the continued presence of less graphic videos on the platform generated a tremendous amount of controversy.

Faced with mounting public pressure, Twitter, Facebook, Microsoft, and YouTube convened on June 26, 2017 to form the Global Internet Forum to Counter Terrorism, or GIFCT. The self-proclaimed goal of this new organization was to make the spaces that these platforms hosted online "hostile to terrorists and violent extremists" and emphasized an approach rooted in technological solutions, research, and knowledge-sharing among corporations.[68] More specifically, these methods included techniques such as the adoption of a "Shared Industry Hash Database" that sought to make flagging content for investigation easier, an outreach program to incorporate smaller companies, and a counter-speech initiative that sought to amplify more positive inclusive speech.[69] Notably, the original iteration of the GIFCT was entirely industry led and funded, making it one of the first true entirely corporate-based organizations dedicated to a unified process of regulation online.

66 Rita Katz, "To Curb Terrorist Propaganda Online, Look to YouTube. No, Really." Security, *Wired,* October 20, 2018, https://www.wired.com/story/to-curb-terrorist-propaganda-online-look-to-youtube-no-really/

67 Zack Beauchamp, "ISIS captured and executed James Foley and Steven Sotloff, two American journalists," *Vox,* November 17, 2015, https://www.vox.com/2018/11/20/17996042/isis-captured-and-executed-james-foley-and-steven-sotloff-two-american-journalists

68 Twitter Public Policy, "Global Internet Forum to Counter Terrorism," June 26, 2017, https://blog.twitter.com/en_us/topics/company/2017/Global-Internet-Forum-to-Counter-Terrorism.html

69 Ibid.

Organizations such as GIFCT and the individual corporations that comprise it possess a substantial edge over the far more cautious law-centric approach when it comes to regulating online content. With direct access to the algorithms and tools that monitor, sort, and display content, these companies are capable of making rapid adjustments to what users can post, see, and share on their platforms. Further, this advanced level of control can be exerted on both a short-term basis to selectively isolate and remove a single piece of content, and a long-term basis to alter sorting and display algorithms to affect a broad range of content. To understand the short-term approach, it is useful to return to the example of how Facebook handled the upload of the Christchurch attack video. In the event of a crisis such as that posed by the Christchurch video, the content moderation team at Facebook initiate a three step process to determine how large of a threat a selective piece of content poses.[70] The first phase in this process is "understand" in which the content is analyzed to determine whether or not it may violate content standards created by Facebook.[71] In the case of the Christchurch video, the video was in clear violation of the "Dangerous Individuals and Organizations policy, which bans "organizations or individuals that proclaim a violent mission or are engaged in violence."[72] This then triggers the second phase, "isolate," in which Facebook moderators move to prevent the spread of content any further on the website.[73] This process is typically undertaken using "hash" technology in which a form of "digital fingerprint" is taken from a video by selecting a specific group of pixels and turning them into a

70 Klonick, "Inside the Team at Facebook That Dealt with the Christchurch Shooting."
71 Ibid.
72 Ibid.
73 Ibid.

traceable numerical identification tag (a hash).[74] This allows any subsequent upload of the video to be quickly identified by the presence of this unique hash and then removed. Finally, once the video has been contained the moderator teams move into the "enforcement" phase in which they continue to monitor to ensure that the banned content doesn't somehow reemerge.[75]

While the three-step approach utilized by Facebook is useful for the rapid isolation and removal of a single piece of content, it isn't particularly useful for a large influx of unwanted media such as the case of the thousands of jihadist videos on YouTube. In these cases, companies adopt a more long term approach intended to alter more generalized patterns in their sorting, display, and removal algorithms. In the YouTube example, Google (YouTube's parent company) made several broader changes to regulate uploaded videos. Among these changes was the addition of A.I. to help identify and classify unwanted content, an increase in human moderators to flag content, demonetization of videos that do not clearly violate policy but contain potentially inflammatory messages, and an expansion of YouTube's pre-existing counter-radicalization programs.[76] These strategies work to tighten YouTube's filters and drive out undesired content from their platform in a more blanket manner. In the case of ISIS related videos, this strategy was largely successful with a substantial observable drop in ISIS content evident between 2017 and 2018.[77]

74 Ibid
75 Ibid.
76 Kent Walker, "Four steps we're taking today to fight terrorism online," *Google,* June 18, 2017, https://www.blog.google/around-the-globe/google-europe/four-steps-were-taking-today-fight-online-terror/
77 Katz, "To Curb Terrorist Propaganda Online, Look to YouTube. No, Really."

Where a corporate-centric approach thus excels is where a law-centric approach largely fails: it has incredible flexibility and is capable of adjusting to newly presented content incredibly quickly. Armed with moderator teams working around the globe and advanced tools for identifying and removing rapidly proliferating content, these corporations maintain the ability to regulate at lightning speed. Yet in returning to both the short-term and long-term approaches previously highlighted, so too does a profound weakness to the corporate-centric approach emerge. In both cases, the process to identify and subsequently act on content is based entirely on standards created by the corporations themselves. Extremely powerful tools such as the hash database utilized by the companies behind GIFCT are entirely opaque to those outside of the companies that use them, meaning "We don't know what companies are feeding into the database, how many false positives there are, or how many users appeal such decisions."[78]

This lack of transparency is incredibly alarming considering just how integral companies involved in an organization like GIFCT are to global communication. YouTube has *two billion users every month (nearly a third of the entire internet)*[79] and yet the majority of the algorithms and methods used to regulate what content is displayed to users is known only within the company itself. While this enables these companies to move quickly in the regulation process, so too does it enable them to do so with little accountability beyond their own internal standards. This in turn leads to the potential for excessive and harmful instances of censorship, a potential that is already being realized in some capacity. According to Google's transparency report, YouTube removed 33 million

78 Jillian C. York, "The Christchurch Call Comes to the UN," *Electronic Frontier Foundation*, September 26, 2019, https://www.eff.org/deeplinks/2019/09/christchurch-call

79 "YouTube by the Numbers," YouTube for Press, YouTube About, accessed March 8, 2020, https://www.youtube.com/about/press/

videos in 2018 (around 90,000 a day), with 73% of those being removed by automated processes before the videos were even available for viewing.[80] These automated processes are a necessity for YouTube to be able to filter through such a high volume of content, yet the way they function has led to some worrisome results. In order to identify prohibited content, YouTube utilizes machine-learning algorithms that are capable of recognizing and flagging instances of prohibited content (such as pornography or graphic violence) by comparing imagery in the pending upload with templates of problematic videos. Yet frequently, this system produces false positives, with content from organizations such as the Syrian Observatory for Human Rights getting caught in this automated net.[81] Lacking a critical degree of nuance, videos such as those depicting government airstrikes on hospitals and medical facilities are automatically flagged and terminated by these automated processes due to their depictions of violence.[82] Similar instances of censorship of conflict documentation have also occurred on channels covering human rights atrocities in Yemen and the Ukraine creating a crisis in which crucial evidence of human rights atrocities is being automatically terminated with no real avenue for recourse.[83]

Thus, similar to a law-centric approach, a corporate-centric approach contains substantial benefits but also suffers from crucial weaknesses. Although it is tempting to turn towards the rapid response capabilities of a corporate-centric approach when confronted by the often glacial speed of law-centric processes, doing so comes at the cost

80 Jillian C. York, "Caught in the Net: The Impact of "Extremist" Speech Regulations on Human Rights Content," *Electronic Frontier Foundation,* May 30, 2019, https://www.eff.org/wp/caught-net-impact-extremist-speech-regulations-human-rights-content

81 Ibid.

82 Ibid.

83 Jillian C. York, "Caught in the Net."

of high rates of censorship and a distinct lack of transparency. Even with the adoption of "advanced machine learning"[84] methods in an attempt to more accurately flag content, it is evident from examples such as the repeated censorship of conflict documentation channels on YouTube that the current corporate-centric approach remains somewhat overzealous in its approach to regulation. To conclude this project, we must then attempt to develop a hybrid model that seeks to adopt strengths from both a corporate-centric and law-centric approach. To imagine what such a model might look like, we must turn to the Christchurch Call.

Part III - Moving Forward: The Christchurch Call and Beyond

On May 15, 2019, two months after the Christchurch tragedy, the New Zealand prime minister Jacinda Arden and French President Emmanuel Macron called upon nations around the world as well as major tech corporations to adopt the Christchurch Call.[85] As described by the official website itself, the Call "outlines collective, voluntary commitments from Governments and online service providers intended to address the issue of terrorist and violent extremist content online and to prevent the abuse of the internet as occurred in and after the Christchurch attacks."[86] Supported by over 60 countries (including the council of Europe and European Commission) and seven major online service providers (including tech giants such as Google, Amazon, and Facebook),

84 Kent Walker, "Four steps we're taking today to fight terrorism online."

85 "Christchurch Call to Eliminate Terrorist & Violent Extremist Content Online," Home, New Zealand Ministry of Foreign Affairs and Trade, accessed March 9, 2020, https://www.christchurchcall.com/call.html

86 "Christchurch Call" The Call.

the Call marks one of the first true unified efforts between civil society and private platforms to engage in responsible regulation of the internet.[87]

The primary structure of the Call is divided into three separate categories outlining the expected commitments from government organizations, private service providers, and finally both of these entities working collectively. Expectations for governments include the strengthening of national inclusiveness and resilience to "counter the drivers of terrorism and violent extremism," encouragement of ethical standards for media outlets, and the ensurement of "effective enforcement" of laws related to the production and dissemination of terrorist content.[88] Expectations for online service providers are somewhat more specific, including the review of content moderation algorithms, strict enforcement of established community standards and terms of service, expansion of transparency through clear publication of what constitutes a violation, and continued development of moderation tools such as hash databases.[89] Finally, expectations for both government and service providers collectively include the acceleration of research and development of technical solutions for more effective moderation and content removal, ensure the cooperation of all parties with law enforcement agencies, support existing academic efforts to understand online extremism, and support smaller underdeveloped platforms in their establishment of effective anti-terrorist moderation systems.[90]

87 "Christchurch Call," Supporters.
88 "Christchurch Call," The Call.
89 Ibid.
90 "Christchurch Call," The Call.

Additionally, the Call pushed for a substantial restructuring and refinement of the previous industry-exclusive Global Internet Forum to Counter Terrorism (GIFCT). Following initiatives outlined by the Call, in September of 2019 the GIFCT announced the introduction of a standardized *Content Incident Protocol* intended "to guide a collaborated response amongst GIFCT members to terrorist attacks," released algorithms related to the hash database utilized by GIFCT members, and published the first GIFCT transparency report.[91] Likewise, although governance of the GIFCT remains within the hands of the industry-led Operating Board, an Independent Advisory Committee and a Multi-stakeholder Forum have been established with the intent of integrating the voices of civil society members and organizations such as advocacy groups and human rights organizations into the GIFCT decision-making process.[92]

While the Christchurch Call represents the creation of a broad, ambitious framework that seeks to unify civil society and private service providers, it lacks specificity in its approach. This lack of specificity has led organizations committed to the defense of civil liberties online such as the Electronic Frontier Foundation (EFF) to additionally advocate for the creation of a concrete collection of principles to guide future attempts at regulation, particularly in the private sphere. From this desire emerged the Santa Clara Principles: Formed on May 7th, 2018 by a small collection of individuals and organizations including the ACLU Foundation of Northern California, the Center for Democracy and Technology, and the Electronic Frontier Foundation itself, the Santa Clara principles outline three concrete methods through which the powerful reach of

91 "Next Steps for GIFCT," Global Internet Forum to Counter Terrorism, September 3, 2019, https://gifct.org/press/next-steps-gifct/
92 Ibid.

corporate regulation can be applied in a manner that maintains the protections created through a law-centric approach. The principles are as follows: "Numbers," in which companies must publish specific numerical data about how many posts are removed and how many accounts are banned from their sites, "Notice," in which companies must clearly outline their content guidelines and provide a specific explanation regarding any content removal or user ban, and "Appeal," in which companies must provide a method for users to appeal any content termination or account suspension.[93] The purpose of these principles is to "provide meaningful due process to impacted speakers and better ensure that the enforcement of their [service provider's] content guidelines is fair, unbiased, proportional, and respectful of users' rights."[94] Additionally, these principles are intended to serve as a minimum baseline for companies rather than a definitive ceiling with the hope that their adoption can serve as a way to further the pursuit of transparency and accountability in the process of online regulation.[95]

Adoption of the Santa Clara Principles by most major tech companies has been slow, but progress has nevertheless been made. According to a May 2019 report by the New America Open Institute of Technology, Facebook, Twitter, and YouTube have taken steps towards the development of recommendations in the "notice" and "appeals" categories of the Principles, but "still fall woefully short when it comes to implementing the recommendations put forth for the "numbers" category."[96] Although all three of these

93 "The Santa Clara Principles: On Transparency and Accountability in Content Moderation," accessed December 13, 2019, https://santaclaraprinciples.org/

94 Ibid.

95 Ibid.

96 "One Year After the Release of the Santa Clara Principles, OTI Continues to Push Tech Companies for Transparency and Accountability Around Content Moderation Practices," Open Institute of Technology, *New America,* May 7, 2019, https://www.newamerica.org/oti/press-releases/one-year-after-release-santa-

major companies have taken the previously unprecedented step of issuing transparency reports related to content moderation practices, the study found that a "significant need for improvement" still existed.[97] At the time of the study, glaring issues such as YouTube's continued failure to provide transparency regarding automated tools for content takedowns and Facebook's lack of a single specific counter for all content removal across the platform demonstrate that these platforms still remain far from fully transparent and continue to wield enormous unchecked regulatory power.[98]

Despite somewhat sluggish adoption the Christchurch Call and Santa Clara principles offer a valuable and mutually complementary framework through which to approach online content moderation in a just and effective manner. Yet as these newly emergent solutions seek to lay the groundwork for creating a hybrid approach between corporate-centric and law-centric methodologies, they reveal that such a hybrid is not necessarily a harmonious entity. Having long maintained independent regulatory power, many private online service providers have been noticeably reluctant to hand over the reins to civil society.[99] Examining the actual adoption of the Call and the Principles only

clara-principles-oti-continues-push-tech-companies-transparency-and-accountability-around-content-moderation-practices/

97 New America, "One Year After the Release of the Santa Clara Principles,"

98 New America, "One Year After the Release of the Santa Clara Principles,"

99 This project largely operates within the conception that the online spaces operated by service providers such as Facebook, Twitter, YouTube, Reddit, etc. constitute true "private" spaces distinct from "public" spaces operated and maintained by civil society. This "private" designation grants a form of protection from First Amendment regulations that only extend to "public" spaces, yet notably this public/private distinction has been challenged in the Supreme Court case *Marsh v. Alabama* (1946). *Marsh* involved the refusal by a privately owned "company town" to allow the distribution of literature by Jehovah's witnesses. The court ruled that such a refusal was in violation of the First Amendment as although the town was technically a "private" entity, it served many of the same functions as a public municipality. Further, the court stated that the more a private entity opens up to the public for their own benefit, the more beholden such an entity is to the statutory and constitutional rights of those who use it (Marsh v. Alabama, 325 U.S. 509 (1968)). In many ways, websites such as Facebook and Twitter have come to resemble the company town in *Marsh,* serving as forums for political debate and even being utilized by the President of the United States as a form of official broadcast network. While many of these websites cling desperately to their

helps to reinforce this. Critics have pointed out how massively influential organizations such as the GIFCT remains governed by industry corporations *only,* how hash databases utilized by these corporations remain largely opaque, and how the crucial terms "terrorism" and "violent extremism" remain up to the discretion of companies themselves.[100]

This is not to say that the Call or the Principles must be disregarded entirely, but rather that we must recognize that even as we forge methods that may *seek* to balance corporate-centric and law-centric approaches, in practice such approaches are comprised of somewhat conflicting principles. My ultimate goal with this project is to demonstrate that there is no singular solution to the complex and dangerous threat posed by white supremacy online and that even while we ultimately strive for an approach that combines the best elements of two starkly different methods such a process is by no means easy or perhaps even truly accomplishable. The internet has long existed as a sort of unregulated frontier for vast global communication, yet as the continually growing web of white supremacist attacks on an international stage demonstrates the days of such a wild frontier may be soon coming to a close.

distinction as "private" spaces, the artificial division between private and public becomes increasingly blurred as these online spaces expand in scope and importance to society.

100 Jillian C. York, "The Christchurch Call: The Good, the Not-So-Good, and the Ugly," Electronic Frontier Foundation, May 16, 2019 https://www.eff.org/deeplinks/2019/05/christchurch-call-good-not-so-good-and-ugly

Bibliography

Abrams v. United States, 250 U.S. 616, (1919)

Anti-Defamation League. "Pepe the Frog." General Hate Symbols. Accessed February 29, 2020. https://www.adl.org/education/references/hate-symbols/pepe-the-frog

Beauchamp, Zack. "ISIS captured and executed James Foley and Steven Sotloff, two American journalists." November 17, 2015. https://www.vox.com/2018/11/20/17996042/isis-captured-and-executed-james-foley-and-steven-sotloff-two-american-journalists

Brandenburg v. Ohio, 395 U.S. 444, (1969).

Butler, Judith. "Limits on Free Speech?" *Academe Blog*, December 7, 2017. https://academeblog.org/2017/12/07/free-expression-or-harassment/

Cai, Weiyi and Landon, Simone. "Attacks by White Extremists Are Growing. So Are Their Connections." *The New York Times,* April 3, 2019. https://www.nytimes.com/interactive/2019/04/03/world/white-extremist-terrorism-christchurch.html

Debs v. United States, 249 U.S. 211, (1919).

"First Amendment — Freedom of Speech — Second Circuit Affirms Threats Conviction in Internet Speech Case — United States v. Turner." Harvard Law Review 127, no. 8 (June 2014): 2585. JSTOR.

Greene, Viveca S. "Deporable' Satire: Alt-Right Memes, White Genocide Tweets, and Redpilling Normies." *Studies in American Humor* 5, no. 1 (2019): 38. Project MUSE.

Global Internet Forum to Counter Terrorism. "Next Steps for GIFCT." September 3,

2019. https://gifct.org/press/next-steps-gifct/

Hammack, Scott. “The Internet Loophole: Why Threatening Speech On-line Requires a Modification of the Courts’ Approach to True Threats and Incitement.” Columbia Journal of Law and Social Problems 36, no. 1 (Fall 2002): 12.

Katz, Rita. “To Curb Terrorist Propaganda Online, Look to YouTube. No, Really.” Security. *Wired*, October 20, 2018. https://www.wired.com/story/to-curb-terrorist-propaganda-online-look-to-youtube-no-really/

Klonick, Kate. “Inside the Team at Facebook That Dealt with the Christchurch Shooting.” *The New Yorker,* April 25, 2019. https://www.newyorker.com/news/news-desk/inside-the-team-at-facebook-that-dealt-with-the-christchurch-shooting

Mackinnon, Catherine. *Only Words* (Cambridge: Harvard University Press, 1993).

Masses Publ'g v. Patten, 244 F. 535, (S.D.N.Y. 1917).

Merriam-Webster.com. "Incite.” 2020. https://www.merriam-webster.com (13 February 2020).

Mill, John Stuart, “Of Individuality, as One of the Elements of Well-Being,” in *On Liberty and Other Writings,* ed. Stefan Collini (Cambridge: University Press, 1989), 57.

New Zealand Ministry of Foreign Affairs and Trade. “Christchurch Call to Eliminate Terrorist & Violent Extremist Content Online.” Home. Accessed March 9, 2020. https://www.christchurchcall.com/call.html

Open Institute of Technology. “One Year After the Release of the Santa Clara Principles,

OTI Continues to Push Tech Companies for Transparency and Accountability Around Content Moderation Practices." New America. May 7, 2019. https://www.newamerica.org/oti/press-releases/one-year-after-release-santa-clara-principles-oti-continues-push-tech-companies-transparency-and-accountability-around-content-moderation-practices/

Oyez. "Whitney v. California." Facts of the Case. Accessed February 16, 2020. https://www.oyez.org/cases/1900-1940/274us357

Roose, Kevin. "A Mass Murder of, and for, the Internet." *The New York Times*, March 15, 2019.https://www.nytimes.com/2019/03/15/technology/facebook-youtube-christchurch-shooting.html

Roose, Kevin. "'Shut the Site Down,' Says the Creator of 8chan, a Megaphone for Gunmen." *The New York Times,* August 4, 2019. https://www.nytimes.com/2019/08/04/technology/8chan-shooting-manifesto.html

Schenck v. United States, 249 U.S. 47, (1919).

Southern Poverty Law Center. "Alt-Right." Accessed February 18, 2020. https://www.splcenter.org/fighting-hate/extremist-files/ideology/alt-right

Taplin, Jonathan. "How to Force 8Chan, Reddit, and Others to Clean Up." *The New York Times*, August 7, 2019. https://www.nytimes.com/2019/08/07/opinion/8chan-reddit-youtube-el-paso.html.

Terminiello v. Chicago, 337 U.S. 1, 4, (1949).

"The Santa Clara Principles: On Transparency and Accountability in Content Moderation." Accessed December 13, 2019. https://santaclaraprinciples.org/

Twitter Public Policy. "Global Internet Forum to Counter Terrorism." Last modified June

26, 2017. https://blog.twitter.com/en_us/topics/company/2017/Global-Internet-Forum-to-Counter-Terrorism.html

U.S. Const. amend. I

Vile, R. John. "Masses Publishing Co. v. Patten (S.D.N.Y) (1917)." *The First Amendment Encyclopedia*, accessed February 16, 2020. https://www.mtsu.edu/first-amendment/article/502/masses-publishing-co-v-patten-s-d-n-y

York, C. Jillian, "Caught in the Net: The Impact of "Extremist" Speech Regulations on Human Rights Content." *Electronic Frontier Foundation*, May 30, 2019. https://www.eff.org/wp/caught-net-impact-extremist-speech-regulations-human-rights-content

York, C. Jillian. "The Christchurch Call Comes to the UN." *Electronic Frontier Foundation*, September 26, 2019. https://www.eff.org/deeplinks/2019/09/christchurch-call

York, C. Jillian."The Christchurch Call: The Good, the Not-So-Good, and the Ugly." *Electronic Frontier Foundation*, May 16, 2019. https://www.eff.org/deeplinks/2019/05/christchurch-call-good-not-so-good-and ugly

Walker, Kent. "Four steps we're taking today to fight terrorism online." Google, June 18, 2017.https://www.blog.google/around-the-globe/google-europe/four-steps-were-taking-today-fight-online-terror/

Whitney v. Cal., 274 U.S. 357, (1927).

Wilson, K. John, "In Opposition to Butler's 'Limits on Free Speech," *Academe Blog*,

December 12, 2017, https://academeblog.org/2017/12/12/in-opposition-to-butlers-limits-on-free-speech/

YouTube About "Youtube by the Numbers." YouTube for Press. Accessed March 8, 2020. https://www.youtube.com/about/press/