LINCOLN MEMORIAL UNIVERSITY
LAW REVIEW

VOLUME 10 SPRING 2023 ISSUE 2

HAS THE "CYBER-MARKETPLACE OF IDEAS" GONE ROGUE?

Matthew J. Blaney[1]

I. INTRODUCTION

The internet and social media have snatched the heartstrings of the world, and what do we have to show for it? For starters, society derives clear and substantial benefits from using social media.[2] Social media allows marginalized individuals to collectively hold powerful individuals accountable.[3] Take, for example, when a police officer used excessive force to murder George Floyd in Minneapolis.[4] Following Floyd's killing, videos of the brutality instantly ricocheted around the world, spurring millions to flood the streets and social media feeds espousing "#BlackLivesMatter."[5] Furthermore, the "#MeToo Movement," a product of social

---

[1] Matthew J. Blaney, J.D. Candidate, Washington University in St. Louis School of Law, 2024; B.A., Criminology, Benedictine College, 2021.

[2] Andrew J. Ceresney et al., *Regulating Harmful Speech on Social Media: The Current Legal Landscape and Policy Proposals*, *in* SOCIAL MEDIA, FREEDOM OF SPEECH AND THE FUTURE OF OUR DEMOCRACY xxiii, xxiii (Lee C. Bollinger & Geoffrey R. Stone eds.,2022).

[3] *Id.*

[4] *See* Valerie Wirtschafter, *How George Floyd changed the online conversation around BLM*, BROOKINGS INST. (June 17, 2021), https://www.brookings.edu/techstream/how-george-floyd-changed-the-online-conversation-around-black-lives-matter/.

[5] *See id.*

media's capacity to connect advocates, has fostered increased accountability for prominent individuals accused of sexual misconduct.[6] These social movements do not stand alone; rather, since its early days, the internet has prompted significant changes in a range of industries—transforming the lives of countless people.[7]

In a 1995 article, Professor Eugene Volokh predicted that emerging technologies would lead to the rapid dissemination of information.[8] Volokh used the term "cheap speech" to refer to the tremendous influx of speech that would be facilitated by new technology because he realized that speakers—both rich and poor—would be able to share their messages widely.[9] Inasmuch as Professor Volokh predicted that technology would make speech more democratic and diverse, he was correct.[10] "By dispersing information sources," Professor Gregory Magarian explained, "cheap speech significantly decentralizes public discourse."[11] Moreover, "[d]ecentralization gives people much greater control over what they say and what they hear than when speech was more expensive."[12]

In theory, a platform where individuals from any race, gender, socioeconomic status, religion, national origin, political ideology, etc., may communicate and collaborate with unprecedented ease is ideal for democracy.[13] But "[w]hat will happen when the KKK becomes able to conveniently send its views to hundreds of thousands of supporters throughout the country?"[14] When Professor Volokh wrote his article in 1995, he warned of the dangers of extremists with access to the media,

---

[6] *See* Seema Shukla et al., #*ME TOO MOVEMENT: INFLUENCE OF SOCIAL MEDIA ENGAGEMENT ON INTENTION TO CONTROL SEXUAL HARASSMENT AGAINST WOMEN*, 12 J. CONTENT, CMTY. & COMMC'N 57, 57 (2020).

[7] *See* Michael Karanicolas, *Squaring the Circle Between Freedom of Expression and Platform Law*, 20 U. PITT. J. TECH. L. POL'Y 177, 210 (2020).

[8] *See* Eugene Volokh, *Cheap Speech and What It Will Do*, 104 YALE L. J. 1805, 1806 (1995).

[9] *See id.* at 1806–07.

[10] *See id.*

[11] Gregory P. Magarian, *How Cheap Speech Underserves and Overheats Democracy*, 54 U.C. DAVIS L. REV. 2455, 2460 (2021).

[12] *Id.*

[13] *See id.* at 2456.

[14] Volokh, *supra* note 8, at 1848.

of falsehoods with an audience in the millions, and of an ill-informed electorate.[15] Nevertheless, to him, the dangers of content regulation were even greater.[16] But is such a position still tenable today?

In recent years, social media has been inundated by a tsunami of hate and xenophobia,[17] as well as a plethora of falsehoods,[18] forging a toxic environment that makes cyberspace appear as if it is falling to its death. Kanye West exclaimed he would soon go "death con 3 On JEWISH PEOPLE";[19] a conservative social media pundit sent a swarm of Twitter users to harass Elliot Page, a transgender actor, with the wrong pronoun and name;[20] and QAnon, a once-fringe conspiracy theory, went mainstream, flooding social media with false information and attempts to increase their ranks.[21]

When approaching the arduous issue of how to handle the flood of hate speech, many scholars have agreed on one thing: The current state of social media content moderation is a disaster.[22] The power of social media platforms to control

---

[15] *See id.* at 1849.

[16] *See id.*

[17] *See* Press Release, Secretary-General, Secretary-General Denounces 'Tsunami' of Xenophobia Unleashed amid COVID-19, Calling for All-Out Effort against Hate Speech, U.N. Press Release SG/SM/20076 (May 8, 2020), https://press.un.org/en/2020/sgsm20076.doc.htm.

[18] *See, e.g.*, Tiffany Hsu, *41 Million Americans are QAnon Believers, Survey Finds*, N.Y. TIMES (Feb. 24, 2022, 12:01 AM), https://www.nytimes.com/live/2020/2020-election-misinformation-distortions.

[19] Stuart A. Thompson, *Kanye West's Posts Land Him in Trouble on Social Media*, N.Y. TIMES, Oct. 10, 2022, at B5.

[20] Amanda Seitz, *LGBTQ Harassment, Slurs Abound on Social Media, Report Says*, ASSOCIATED PRESS (July 13, 2022), https://apnews.com/article/technology-gay-rights-elliot-page-4e2208f50ba038086f55f3021a7af397.

[21] Kevin Roose, *What Is QAnon, the Viral Pro-Trump Conspiracy Theory?*, N.Y. TIMES, Aug. 19, 2020, at B1.

[22] *See* Nadine Strossen, *The Paradox of Free Speech in the Digital World: First Amendment Friendly Proposals for Promoting User Agency*, 61 WASHBURN L.J. 1, 43 (2021) ("The unprecedented, constitutionally unconstrained power of dominant platforms to restrict free speech and user agency presents a major challenge."); Brittany Doyle, *Self-Regulation Is No Regulation – the Case for Government Oversight of Social Media Platforms*, 32 IND. INT'L & COMP. L. REV. 97, 129–30 (2022) ("Even

global, public discourse is so enormous that their level of influence rivals—and in some cases even exceeds—that of nation-states.[23]

Some cases suggest that Big Tech's power to regulate online speech is reasonable. For example, Stephen Ayres—a man who pleaded guilty to breaching the Capitol building on January 6, 2021—admitted that President Trump's social media posts motivated rioters to storm the Capitol that day.[24] On the other hand, however, those who are outspoken about white supremacy and racism have reported finding their content removed or taken down for violating community guidelines.[25] For many activists, social media is their primary way to reach citizens, and their messages may, at times, be removed without any option to appeal.[26] To make the matter more concerning, private platforms that censor speech are non-state actors exercising their own First Amendment rights; thus, such censorship does not violate the First Amendment.[27]

Mark Zuckerberg said, 'Facebook should not make so many important decisions about free expression and safety on our own.'"); Nadine Strossen, *United Nations Free Speech Standards as the Global Benchmark for Online Platforms' Hate Speech Policies*, 29 MICH. ST. INT'L. L. REV. 307, 361 (2021) ("Given the Platforms' preeminence as venues for free speech and for the political discourse that is the lifeblood of democratic self-government, their unprecedented and unconstrained censorial powers threaten essential foundations of liberal democracies: free speech, equality, due process of law, and the rule of law.").

[23] *See* Karanicolas, *supra* note 7, at 210.

[24] *See* Tom Dreisbach, *How Trump's 'will be wild!' tweet drew rioters to the Capitol on Jan. 6*, NPR (July 13, 2022, 3:42 PM), https://www.npr.org/2022/07/13/1111341161/how-trumps-will-be-wild-tweet-drew-rioters-to-the-capitol-on-jan-6.

[25] *See* Janice Gassam Asare, *Social Media Continues To Amplify White Supremacy And Suppress Anti-Racism*, FORBES (Jan. 8, 2021, 8:43 PM), https://www.forbes.com/sites/janicegassam/2021/01/08/social-media-continues-to-amplify-white-supremacy-and-suppress-anti-racism/?sh=4c4b0d104170.

[26] *See id.*

[27] *See* Nadine Strossen, *The Paradox of Free Speech in the Digital World: First Amendment Friendly Proposals for Promoting User Agency*, 61 WASHBURN L.J. 1, 3 (2021).

The answer to this Free Speech dilemma, some scholars argue, is that we must curb non-governmental censorship.[28] As is, modern technology allows individuals to pepper vulnerable targets with social media notifications, creating an information overload that overwhelms users and leaves them susceptible to bad actors seeking to spread hate speech.[29] With this in mind, is it truly the wisest idea to open the floodgates and let internet users create and share any content they want within a matter of seconds? Absolute protection of online speech indeed seems to promote the democratic exchange of ideas, but what are we to think when online hate speech translates to real-world harm?

This note makes a case for regulating online hate speech, but also seeks to answer fundamental questions as to why we protect speech in the first space. Is online speech fundamentally different from offline speech and, therefore, possibly entitled to less protection? Does First Amendment theory justify protecting online hate speech when the framers wrote the First Amendment nearly two centuries before the invention of the internet?

In answering these questions, Part I of this article will give a general background on current hate speech jurisprudence, including a discussion of the categories of speech that do and do not receive protection.

Part II addresses two primary concerns: First, the internet allows hate speech to be disseminated with unprecedented ease; and second, online speech is inherently different from offline speech. This note argues that both of these concerns demand that we, as a society, re-examine the reasons we protect speech in the first place.

Part III of this note will analyze the theoretical basis for First Amendment rights. Subsequently, this note argues that the First Amendment theory no longer justifies protecting online hate speech.

Lastly, Part IV will propose a compromise in the debate over online hate speech regulation. This note argues that section 230 of the Communications Decency Act should be protected so platforms will retain their First Amendment rights to choose what content shall appear on their pages. Additionally, for

---

[28] *See id.*

[29] *See* Yitzchak Besser, *Web of Lies: Hate Speech, Pseudonyms, the Internet, Impersonator Trolls, and Fake Jews in the Era of Fake News*, 17 OHIO ST. TECH. L.J. 233, 244 (2021).

platforms that do choose to moderate online hate speech, a method for doing so—which promotes transparency, procedural rights, oversight, and limiting institutional bias in making content moderation decisions—will be analyzed.

II. BACKGROUND

American culture has changed dramatically in the forty-five years since the Illinois Supreme Court held that the First Amendment protected the right of the American Nazi Party to display the swastika during a march through the Jewish community of Skokie, Illinois.[30] Despite these changes, American jurisprudence continues to protect "ideas that the overwhelming majority of people might find distasteful or discomforting."[31] That is not to say that Americans are free to proclaim any thought that passes through their minds without consequence. But unless their speech constitutes fighting words,[32] obscenity,[33] defamation,[34] or incitement to violence,[35] it likely is protected from infringement by the government.

Notwithstanding, the rapid growth of the internet may demand that Americans re-examine the kinds of speech that promote and inhibit democracy. But first, the doctrinal framework underpinning First Amendment jurisprudence must be surveyed.

A. A HISTORY OF HATE SPEECH JURISPRUDENCE

1. Klan Rallies, Nazi Marches, & Burning Crosses: Just How Flexible is Freedom of Speech?

In *United States v. Schwimmer*, Justice Holmes— writing in dissent—penned a phrase that would shape the free speech debate for years to come:

---

[30] *See generally* Skokie v. Nat'l Socialist Party, 373 N.E.2d 21 (Ill. 1978).
[31] Virginia v. Black, 538 U.S. 343, 358 (2003) (quoting Abrams v. United States, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting)).
[32] *See* Chaplinsky v. New Hampshire, 315 U.S. 568 (1942).
[33] *See* Miller v. California, 413 U.S. 15 (1973).
[34] *See* New York Times Co. v. Sullivan, 376 U.S. 254 (1964).
[35] *See* Brandenburg v. Ohio, 395 U.S. 444 (1969).

> [I]f there is any principle of
> the Constitution that
> more imperatively calls for attachment than any
> other it is the principle of free thought—not free
> thought for those who agree with us but
> freedom for the thought that we hate.[36]

Forty years later, Justice Holmes's statement that the First Amendment should protect the "thought that we hate" reverberated in *Brandenburg v. Ohio*.[37] In *Brandenburg*, the Court convicted a leader of a Ku Klux Klan group under Ohio's criminal syndicalism statute[38] for threatening violence "if our President, our Congress, [and] our Supreme Court, continue[] to suppress the white, Caucasian race. . . ."[39] Ruling in favor of Brandenburg, the Court rejected the proposition that the government can forbid "mere advocacy not distinguished from incitement or imminent lawless action."[40] Instead, the Court held that, in order to be restricted, advocacy of the use of force or violence must be "directed to inciting or producing imminent lawless action" and "likely to incite or produce such action."[41]

As socially reprehensible as wearing Klan regalia, burning crosses, and uttering derogatory comments about black and Jewish Americans may be, the Court nevertheless deemed the statute unconstitutional because it failed to distinguish between mere abstract teaching and preparing a group for violent action.[42] But the Supreme Court was not alone

---

[36] United States v. Schwimmer, 279 U.S. 644, 654–55 (1929) (Holmes, J., dissenting).
[37] Brandenburg v. Ohio, 395 U.S. 444 (1969).
[38] *See id.* at 448 ("The Act punishes persons who 'advocate or teach the duty, necessity, or propriety' of violence 'as a means of accomplishing industrial or political reform'; or who publish or circulate or display any book or paper containing such advocacy; or who 'justify' the commission of violent acts 'with intent to exemplify, spread or advocate the propriety of the doctrines of criminal syndicalism'; or who 'voluntarily assemble' with a group formed 'to teach or advocate the doctrines of criminal syndicalism.'").
[39] *Id.* at 446.
[40] *See id.* at 448–49.
[41] *Id.* at 447.
[42] *See id.* at 448–49.

in its decision to protect the First Amendment rights of a group notorious for advocating hatred of minorities.

Skokie, Illinois, was a quiet suburb nestled immediately north of the City of Chicago and home to 40,000 Jewish residents[43]—comprising over half of the city's population. Life for Skokie's Jewish community would not remain peaceful for long, however. Sporting swastikas on their clothing, Nazis planned to descend on Skokie's streets, displaying their swastikas on their clothing and championing such slogans as "Free Speech for White America."[44] Although the neo-Nazi demonstrators committed to not making derogatory comments, distributing handbills, or blocking traffic,[45] this did not set the minds of Skokie residents at ease.

Among the 5,000 to 7,000 Holocaust survivors in Skokie, one resident testified that the Jewish community felt that the neo-Nazis' message was clear: "the Nazi threat is not over, [the Holocaust] can happen again."[46] In spite of the neo-Nazis' obviously threatening message, the Illinois Supreme Court held that the use of the swastika is symbolic speech protected by the First Amendment.[47] Adhering to U.S. Supreme Court precedent, the court declared that the expression of ideas might not be prohibited because they are offensive to some hearers.[48]

The court did not stop there, however. Faced with the heavy burden of justifying the imposition of a prior restraint on the neo-Nazis' freedom of speech, the village of Skokie sought to employ the fighting words doctrine.[49] The fighting words doctrine—articulated in *Chaplinsky v. New Hampshire*—permits punishment of extremely hostile, personal communications likely to cause an immediate physical response by the person to whom they are directed.[50] And although Skokie argued that the display of the swastika—directed to a community containing

---

[43] *See* Paul Delaney, *Jews in a Chicago Suburb Brace For July 4 March by Nazi Party*, N.Y. TIMES, June 24, 1977, at 8.
[44] *See* Skokie v. Nat'l Socialist Party, 373 N.E.2d 21, 22 (Ill. 1978).
[45] *Id.*
[46] *Id.*
[47] *See id.* at 25.
[48] *Id.* at 23 (citing Bachellar v. Maryland, 397 U.S. 564, 567 (1970)).
[49] *See id*.
[50] *See id.* (citing Chaplinsky v. New Hampshire, 315 U.S. 568, 573 (1942)).

thousands of Holocaust survivors—constituted fighting words, the court disagreed.[51]

Roughly fourteen years later, the U.S. Supreme Court heard an unconventional case involving a Minnesota teenager. In *R.A.V. v. St. Paul*, R.A.V.—a juvenile—along with several other teenagers, assembled a wooden cross and ignited it inside the fenced yard of a black family residing across the street from where R.A.V. was staying.[52] Contrary to the cases discussed above, the First Amendment issue in *R.A.V.* sparked from the ensuing prosecution of R.A.V. under a city ordinance. St. Paul, Minnesota's "Bias-Motivated Crime Ordinance" provided that:

> Whoever places on public or private property a symbol, object, appellation, characterization or graffiti, including, but not limited to, a burning cross or Nazi swastika, which one knows or has reasonable grounds to know arouses anger, alarm or resentment in others on the basis of race, color, creed, religion or gender commits disorderly conduct and shall be guilty of a misdemeanor.[53]

Overbroad in nature, the Minnesota Supreme Court revised the ordinance in an attempt to salvage its constitutionality—limiting the reach of the ordinance to fighting words.[54] Even so, the ordinance, as adjusted, contained a fatal flaw: It prohibited only those fighting words based on race, color, gender, and religion.[55] And unsurprisingly, this led the Supreme Court to strike down the St. Paul ordinance.[56] According to the Court, the ordinance would not cover the use of fighting words associated with homosexuality, political affiliation, etc., and the First Amendment does not allow St. Paul to impose special prohibitions on speakers with disfavored viewpoints.[57]

---

[51] *See id.* at 24.
[52] R.A.V. v. St. Paul, 505 U.S. 377, 379 (1992).
[53] *Id.* at 380 (quoting ST. PAUL, MINN., LEGIS. CODE § 292.02 (1990)).
[54] *See* Akhil Reed Amar, Comment, *The Case of the Missing Amendments: R.A.V. v. City of St. Paul*, 106 HARV. L.REV. 124, 127 (1992).
[55] *See id.*
[56] *See id.*
[57] *R.A.V.*, 505 U.S. at 391.

2. When Is Hate Speech Not Protected?

The protection of non-violent neo-Nazi marchers and Klan meetings may make First Amendment jurisprudence appear as if it ignores the regulation of hate speech. But there are several instances in which hate speech does not receive the same protection as unpopular views that are otherwise legal.

a. True Threats

Not long after the turn of the twenty-first century, the Court in *Virginia v. Black*[58] tackled an issue strikingly similar to the issue deciphered in *R.A.V*. In *Black*, respondents were convicted of violating a Virginia law that made it a felony "for any person …, with the intent of intimidating any person or group ..., to burn ... a cross on the property of another, a highway or other public place. . . ."[59] However, unlike the *R.A.V.* Court, the *Black* Court held that the statute banning cross burning with the intent to intimidate was constitutional.[60] In reaching its decision, the Court explained that the First Amendment permits a State to ban "true threats."[61] And cross burning, which historically has been used to intimidate and instill fear in its victims, is a true threat since the speaker means to communicate a serious expression of an intent to commit an act of unlawful violence to a particular individual or group of individuals.[62]

---

[58] Virginia v. Black, 538 U.S. 343 (2003).

[59] *Id.* at 348. The statute also contained the following provision: "Any such burning ... shall be prima facie evidence of an intent to intimidate a person or group." *Id.* The Court held that this particular provision was unconstitutional, even though the first section of the statute was constitutional. *See id.* at 343–45.

[60] *Id.* at 343.

[61] *Id.* at 359 ("'True threats' encompass those statements where the speaker means to communicate a serious expression of an intent to commit an act of unlawful violence to a particular individual or group of individuals.").

[62] *See id.* at 359–60.

b. Speech in the Private Sector

Aside from the cross-burning controversy of the *Black* case, many people are surprised to learn that the First Amendment restricts only governmental action—meaning that private sector entities generally are allowed to restrict hate speech within their realm of control.[63] And, in reality, private entities certainly do regulate hate speech. Facebook, a private platform, removed more than seven million instances of hate speech in the third quarter of 2019 alone.[64]

c. Time, Place, & Manner Restrictions

Additionally, the government may restrict hate speech if the regulation is a "time, place, and manner" restriction.[65] Courts often will approve of restrictions that are justified without reference to the content of the speech, that leave open ample alternative channels for communicating regulated information, and that serve an important governmental interest.[66] For example, in *Clark v. Community for Creative Non-Violence*, the Court upheld a National Park Service regulation that prohibited camping in certain parks.[67]

In *Clark*, demonstrators descended on Lafayette Park and the National Mall with the intent to camp there to draw attention to the plight of homelessness.[68] Because overnight sleeping in a national park is considered expressive conduct subject to First Amendment examination,[69] at issue was whether the government impermissibly restricted the demonstrators' constitutional rights. The Court held that the regulation was permissible because (1) it was not applied based

---

[63] *See* NADINE STROSSEN, HATE: WHY WE SHOULD RESIST IT WITH FREE SPEECH, NOT CENSORSHIP 57 (1st ed. 2018).

[64] Billy Perrigo, *Facebook Says It's Removing More Hate Speech Than Ever Before. But There's a Catch*, TIME (Nov. 27, 2019, 4:42 AM), https://time.com/5739688/facebook-hate-speech-languages/.

[65] *See* STROSSEN, *supra* note 63, at 57.

[66] *See* Heffron v. Int'l Soc. for Krishna Consciousness, 452 U.S. 640, 647–48 (1981) (quoting Virginia State Bd. of Pharmacy v. Virginia Citizens Consumer Council, Inc., 425 U.S. 748, 771 (1976)).

[67] *See generally* Clark v. Cmty. for Creative Non-Violence, 468 U.S. 288 (1984).

[68] *See id.*

[69] *See id.* at 293.

on disagreement with the message presented; (2) the government had a substantial interest in keeping parks in the heart of our nation's capital in attractive and intact condition; (3) permitting camping would be detrimental to those conditions; and (4) there are methods for highlighting the plight of homelessness other than camping.[70]

Nadine Strossen, author of arguably the most complete, modern work spelling out the limits of hate speech regulation, laid out a prime example of when the government may regulate hate speech with a time, place, and manner restriction. According to Strossen, if a public university sought to ban signs exhibiting hate speech—concerned that the signs may disturb the tranquility of the dormitory—then it could simply ban all signs in the dormitories' shared spaces.[71]

d. Various Government Facilities

Beyond time, place, and manner restrictions, the government may regulate hate speech in particular government facilities—such as prisons and schools.[72] Regarding the latter, case law has addressed the issue of whether students may sport one of the most controversial symbols in American history: the Confederate flag.

Students do not shed their First Amendment rights at the schoolhouse doors;[73] nevertheless, the rights of public school students are not automatically coextensive with the rights of adults in other settings.[74] For example, a student's conduct at school is not protected by freedom of speech when it "materially disrupts classwork or involves substantial disorder or invasion of rights of others."[75]

Candice Hardwick, a student in a small South Carolina town, wore articles of clothing displaying the Confederate flag on several occasions.[76] Candice's school—Latta High School—

---

[70] *See id* at 288.
[71] *See* STROSSEN, *supra* note 63, at 57–58.
[72] *See id*. at 58.
[73] Tinker v. Des Moines Indep. Cmty. Sch. Dist., 393 U.S. 503, 506 (1969).
[74] Bethel Sch. Dist. v. Fraser, 478 U.S. 675, 682 (1986).
[75] *See Tinker*, 393 U.S. at 515.
[76] *See* Hardwick v. Heyward, 711 F.3d 426, 430–32 (4th Cir. 2013).

had a history of fierce racial tension,[77] and Candice's clothing only reinvigorated past conflicts, leading to an in-class controversy surrounding the flag where the teacher had to calm the class down.[78] Accordingly, the Fourth Circuit held that, by prohibiting Candice from wearing Confederate flag t-shirts, Latta school officials did not violate her First Amendment rights.[79]

e. Fighting Words

Off school grounds, courts permit restrictions on another form of disruptive speech that tends to provoke anger in the listener: fighting words. Convicted under a New Hampshire statute proscribing the utterance of "any offensive, derisive or annoying" word or name to a person lawfully in a public place, Walter Chaplinsky challenged the law's constitutionality for restraining his freedom of speech.[80] Rendering a decision that serves as the foundation of the fighting words doctrine to this day, the Court concluded that "[a]rgument is unnecessary to demonstrate that the appellations 'damn racketeer' and 'damn Fascist' are epithets likely to provoke the average person to retaliation, and thereby cause a breach of the peace."[81] As a result, Chaplinsky's speech received no constitutional protection.[82]

Chaplinsky's speech may not be considered hate speech, but, generally, fighting words may resemble hate speech. Take, for example, a member of the Ku Klux Klan uttering a racial epithet at a member of the Black Lives Matter

---

[77] "[I]n the mid–1980s, a white student and an African–American student attended the prom together, causing 'small groups of whites and blacks ... to stir up trouble,' which included white students wearing Confederate flag apparel and African–American students wearing Malcolm X apparel." *Id.* at 432 (citation omitted). "[I]n the mid–1990s, two high school students burned one of the historic African–American churches in the area." *Id.*

[78] *Id.* at 432–33.

[79] *Id.* at 444; *see also* Melton v. Young, 465 F.2d 1332 (6th Cir. 1972) (holding that a school did not violate a student's First Amendment rights by suspending him for wearing clothing sporting the Confederate flag).

[80] *See* Chaplinsky v. New Hampshire, 315 U.S. 568, 569 (1942).

[81] *Id.* at 574.

[82] *See id.*

movement during a face-to-face interaction; such speech presumably could be punished due to its provocative nature.[83]

f. Harassment

Finally, in certain circumstances, hate speech may be punished as harassment.[84] A prime example of when hate speech may be labeled as such arose at Stanford University. While a female African American student made her way across the Stanford campus, a group of white male students followed her shouting, "I've never tried a [N-word] before."[85]

III. HATE SPEECH ONLINE: IS REGULATION LONG OVERDUE?

A. DISTRIBUTING DISCRIMINATION WITH EASE

The days when a citizen had to trek their way to a town-hall meeting in order to have their deepest political thoughts heard are over. There is no need to attend to such an arduous task when you could, in theory, express your thoughts to 36.8 million people in a matter of seconds—solely by leaving a comment on President Biden's public Twitter page.[86] Furthermore, it takes only a matter of seconds to jump over to Elon Musk's Twitter page, where the audience could be as large as 126.5 million people from all over the world.[87] Additionally, social media users can speak their minds behind a veil of anonymity, with no one observing them at the moment they comment.

The shocking reality of how rapidly one can communicate certainly has been realized by social media users across the globe. Politicians use social media to build followings and communicate with constituents;[88] in 2014, the ALS ice

---

[83] *See* STROSSEN, *supra* note 63, at 64.
[84] *Id.*
[85] *See id.* at 65.
[86] *See* @JoeBiden, TWITTER, https://twitter.com/joebiden (last visited Jan. 19, 2023).
[87] *See* @elonmusk, TWITTER, https://twitter.com/elonmusk?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor (last visited Jan. 19, 2023).
[88] *See, e.g.*, Max Benwell, 🔥😂✊: *How Alexandria Ocasio-Cortez beat everyone at Twitter in nine tweets*, THE GUARDIAN (Feb. 12, 2019, 6:00

bucket challenge filled the internet with videos of awkward teenagers dumping ice cold water on themselves to raise awareness of ALS, ultimately leading to $115 million in donations;[89] and, possibly most important of all, most people in the United States now have the ability to make their ideas broadly available.[90]

Social media assuredly has benefited society, especially those individuals and groups previously denied access to traditional forums for democratic discussion. But, currently, the dark side of social media presents a grave threat to the American way of life. The mass shooters of a Pittsburgh synagogue, Buffalo supermarket, and El Paso Walmart all expressed their hateful, violent intentions on social media.[91] In 2017, the organizers of the "Unite the Right" rally—where a car crashed into civilians, killing one and injuring more than a dozen[92]—used the chat app Discord to plan the "Great Replacement Theory"[93] motivated event.[94] Extremism on social

AM), https://www.theguardian.com/us-news/2019/feb/12/alexandria-ocasio-cortez-twitter-social-media.

[89] *See A Guide to Activism in the Digital Age*, MARYVILLE UNIV., https://online.maryville.edu/blog/a-guide-to-social-media-activism/ (last visited Feb. 5, 2023).

[90] *See* Magarian, *supra* note 11, at 2456.

[91] *See* Amanda Seitz, *White supremacists are riling up thousands on social media*, PBS NEWS HOUR (June 10, 2022, 12:07 PM), https://www.pbs.org/newshour/politics/white-supremacists-are-riling-up-thousands-on-social-media.

[92] *See Unrest in Virginia*, TIME, https://time.com/charlottesville-white-nationalist-rally-clashes/ (last visited Jan. 19, 2023).

[93] "In short, the 'great replacement' is a conspiracy theory that states that nonwhite individuals are being brought into the United States and other Western countries to 'replace' white voters to achieve a political agenda." Dustin Jones, *What is the 'great replacement' and how is it tied to the Buffalo shooting suspect?*, NPR (May 16, 2022, 12:35 AM), https://www.npr.org/2022/05/16/1099034094/what-is-the-great-replacement-theory.

[94] *See* Press Release, U.S. Sen. Dick Durbin, Durbin Presses Social Media Platforms on their Roles in Promoting Extremist Content (May 19, 2022), https://www.judiciary.senate.gov/press/dem/releases/durbin-presses-social-media-platforms-on-their-roles-in-promoting-extremist-content.

platforms not only continues to flourish,[95] but there is reason to believe that social media actually intensifies old conflicts and sparks new ones.[96]

Although not entirely clear, the root of the problem, at least partly, stems from one major issue: Private sector censorship is not restricted by the Constitution—on the contrary, it protects it.[97] So, while hate groups, teenagers, and other ordinary citizens are free to post whatever content they please, social media platforms are free to stand by and watch the situation escalate. In reality, platforms such as Facebook do not passively allow the chaos to pan out. Rather, believing that people feel freer to connect and use their voices online when they are not "attacked," Facebook prohibits hate speech on its platform.[98] And although social media companies may not violate the First Amendment when blocking users from their platforms, the fact remains that some form of speech suppression exists.

In addition to her strong argument for protecting offline hate speech, Professor Nadine Strossen puts forth a compelling rationale for allowing uninhibited speech online. According to Strossen, no matter how strictly courts forbid government censorship, society will never enjoy freedom in the digital world unless it combats platform censorship and its corresponding "cancel culture."[99] Additionally, she correctly notes that complaints about the unfettered power of social media platforms have come from all ends of the political spectrum—whether it be President Trump or the Black Lives Matter movement.[100]

---

[95] *See* Ken Dilanian, *FBI, DHS and social media firms like Meta, TikTok aren't adequately addressing threat of domestic extremists, Senate report says*, NBC NEWS (Nov. 16, 2022, 2:00 PM), https://www.nbcnews.com/politics/national-security/fbi-dhs-meta-tiktok-threat-domestic-extremists-rcna57458.

[96] *See* Caitlin Elsaesser, *How Social Media Turns Online Arguments Between Teens Into Real-World Violence*, UCONN TODAY (Apr. 29, 2021), https://today.uconn.edu/2021/04/how-social-media-turns-online-arguments-between-teens-into-real-world-violence-2/#.

[97] Strossen, *supra* note 27, at 4.

[98] *See Hate Speech*, META, https://transparency.fb.com/policies/community-standards/hate-speech/ (last visited Feb. 5, 2023).

[99] Strossen, *supra* note 27, at 4.

[100] *See id.* at 10.

Furthermore, Strossen argues that the definition of hate speech is inherently subjective; what one person considers to be hate speech, another may consider to be "neutral or even loving."[101] For example, members of the Westboro Baptist Church, by directing slurs at LGBTQ individuals in an effort to get them to renounce their sexual orientation or identities, believe they are saving those individuals from "eternal damnation."[102] Thus, Church members believe they are "doing the loving work necessary to save these individuals[]," even though an outsider to the Church may well construe their message as hate speech.[103] Consider also the Blue Lives Matters movement. While some individuals construe the messages conveyed by those advocates as hate speech, many members of the movement believe they are communicating love for police officers.[104]

Private social media companies arbitrarily enforcing inherently vague hate speech codes is uncalled for. And, unfortunately, companies attempting to enforce these policies sometimes suppress the opinions of well-meaning individuals—including those who may be attempting to combat hate speech themselves. But when a problem presents itself, are we just supposed to throw in the towel and allow unfettered speech online? The answer to that question is undoubtedly no, but in order to recognize why, one must recognize the inherent difference between online and offline speech.

B. ONLINE SPEECH IS INHERENTLY DIFFERENT

The unique nature of online speech raises an important concern: If we follow current First Amendment principles, an unlimited amount of hate speech likely would be permitted online.

By way of hypothetical: Say person A and person B are standing in a public area outside the United States Capitol Building, engaging in a debate over whether the Confederate flag truly stands for southern heritage or is merely a racist relic of our nation's past. A, infuriated that B would insult his

---

101 *See id.* at 13.
102 *See id.*
103 *See id.* at 14.
104 *See id.*

southern heritage, runs to his car, grabs a wooden cross, returns to B, lights the cross on fire, and hurls a racial epithet at B—a black citizen.[105]

The threat posed by directing a burning cross and racial epithet at a black citizen more than likely would arouse great fear, and possibly grave anger, in the citizen. As such, there is a good chance that this would be considered a "true threat" or "fighting words"—rendering A's speech unprotected under the First Amendment.[106]

But what if instead of threatening B at the Capitol Building, A tweeted at B a string of racial epithets, a gif of a burning cross, and a message stating that A planned to immediately set B's house on fire?

Assuming First Amendment principles apply to Twitter as they would in front of the Capitol, there is an issue here: A and B are online. Consequently, they are not physically present together, and even if the comment is "directed to inciting or producing imminent lawless action,"[107] it is hardly likely to do so since A and B have all of cyberspace between them. For that same reason, it is highly unlikely that A's Tweet will be unprotected incitement since no immediate physical response will be warranted on B's part. Nor will this likely be considered

---

[105] This fact pattern is not totally absurd: only two years ago, a man referred to as the "QAnon Shaman," wearing face paint and a headdress, broke into the Capitol with other rioters—including a man who smoked marijuana inside the building. *See* Hannah Rabinowitz & Katelyn Polantz, *'QAnon Shaman' Jacob Chansley sentenced to 41 months in prison for role in US Capitol riot*, CNN (Nov. 17, 2021, 1:37 PM), https://www.cnn.com/2021/11/17/politics/jacob-chansley-qanon-shaman-january-6-sentencing/index.html; *see also* *'Doobie Smoker' sentenced: Man seen lighting a joint during Capitol riots gets probation*, 13NEWS NOW (Mar. 4, 2022, 12:42 PM), https://www.13newsnow.com/article/news/crime/doobie-smoker-capitol-rioter-sentencing-probation/291-e2d888e2-7ebb-4153-b260-b46c333dd7c0#:~:text=WASHINGTON%20%E2%80%94%20A%20man%20who%20was,%241%2C000%20and%20a%20%24500%20restitution.

[106] *See* Chaplinsky v. New Hampshire, 315 U.S. 568 (1942) (holding that fighting words are not protected by the First Amendment); *see also* STROSSEN, *supra* note 63, at 60 (explaining that the government may constitutionally punish "true threats").

[107] Brandenburg v. Ohio, 395 U.S. 444, 447 (1969).

a "true threat" as recognized by the *Black* Court[108]—again, because of the sheer impossibility of any immediate harm to B.

Is society to allow the unfettered communication of threatening messages online merely because the current First Amendment doctrine—not originally created to apply to the internet—does not prohibit them? Courts should answer this question with a resounding "no." Speech that would be unprotected offline may find refuge online. But before discussing how this issue should be addressed, this article will first examine the theoretical foundation of the First Amendment and argue that online hate speech does not further its purposes.

IV. THE FIRST AMENDMENT'S THEORETICAL UNDERPINNING

A. TRUTH

At the root of the "marketplace of ideas" are the works of two philosophers: John Milton and John Stuart Mill. Milton's Book, *Areopagitica*—possibly the most imaginative and densely suggestive defense of free speech—has proven to be a foundation for First Amendment jurisprudence.[109] In his famous protest of a licensing scheme for books, Milton wrote:

> And though all the winds of doctrine were let loose to play upon the earth, so truth be in the field, we do injuriously by licensing and prohibiting, to misdoubt her strength. Let her and Falsehood grapple; who ever knew Truth put to the worse, in a free and open encounter.[110]

Two centuries later, in what is considered one of the greatest defenses of free speech, John Stuart Mill echoed a similar sentiment in *On Liberty*.[111] Mill's central thesis was that the suppression of opinion is wrong, regardless of whether the

---

[108] *See* Virginia v. Black, 538 U.S. 343, 359–60 (2003).

[109] *See* Vincent A. Blasi, *A Reader's Guide to John Milton's Areopagitica, the Foundational Essay of the First Amendment Tradition*, 2017 SUP. CT. REV. 273, 273 (2017).

[110] *Id.* at 311 (quoting JOHN MILTON, AREOPAGITICA 746 (1644)).

[111] JOHN STUART MILL, ON LIBERTY (Elizabeth Rapaport ed., Hackett Publishing Co. 1978).

opinion expressed is true.[112] "If [the opinion] is false, society is denied the fuller understanding of truth which comes from its conflict with errors."[113] And "when the received opinion is part truth and part error, society can know the whole truth only by allowing the airing of competing views."[114]

One cannot underplay the influence of Milton and Mill today. Members of the Supreme Court echoed their philosophies on at least two occasions, leading people to refer to this concept as the "marketplace of ideas."[115] When hate speech contains elements that tip the scales of justice in favor of labeling it as "of such slight social value as a step to truth that any benefit that may be derived from [it] is clearly outweighed by the social interest in order and morality,"[116] then it may well be forbidden from the marketplace of ideas. However, absent any extenuating circumstances, hate speech probably will be allowed to flourish in the search for truth.[117]

---

[112] *See* GERALD GUNTHER & KATHLEEN M. SULLIVAN, CONSTITUTIONAL LAW 1025 (13th ed. 1997).

[113] *Id.* (quoting JOHN STUART MILL, ON LIBERTY (Elizabeth Rapaport ed., Hackett Publishing Co. 1978).

[114] *Id.*

[115] *See* Whitney v. California, 274 U.S. 357, 375 (1927) (Brandeis, J., concurring) ("[F]reedom to think as you will and to speak as you think are means indispensable to the discovery and spread of political truth; that without free speech and assembly discussion would be futile; that with them, discussion affords ordinarily adequate protection against the dissemination of noxious doctrine; that the greatest menace to freedom is an inert people; that public discussion is a political duty; and that this should be a fundamental principle of the American government."); *see also* Abrams v. United States, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting) ("[W]hen men have realized that time has upset many fighting faiths, they may come to believe even more than they believe the very foundations of their own conduct that the ultimate good desired is better reached by free trade in ideas—that the best test of truth is the power of the thought to get itself accepted in the competition of the market, and that truth is the only ground upon which their wishes safely can be carried out.").

[116] Chaplinsky v. State of New Hampshire, 315 U.S. 568, 571–72 (1942).

[117] *See, e.g.*, R.A.V. v. St. Paul, 505 U.S. 377, 396 (1992) (stating that, while burning a cross in somebody's yard is reprehensible, the First Amendment should not be added to the fire).

B. DEMOCRATIC SELF-GOVERNMENT

Underlying the Democratic Self-Government justification for freedom of speech is the idea that speech critical of the government's performance ought to be permitted and possibly even encouraged. Alexander Meiklejohn, a leading proponent of this theory, argued that public speech—speech on public issues affecting self-government—must be wholly immune from regulation, while speech regarding private matters should be entitled to less protection.[118]

Meiklejohn's perspective has gained considerable traction among both the Supreme Court and various legal scholars. Justice Black, writing for the majority in *Mills v. Alabama*, maintained that "[w]hatever differences may exist about interpretations of the First Amendment, there is practically universal agreement that a major purpose of that Amendment was to protect the free discussion of governmental affairs."[119] Nevertheless, this position has faced considerable opposition as well.[120]

C. AUTONOMY

The final theoretical justification put forward in defense of free speech rests on valuing individual liberty, autonomy, and self-fulfillment. Justice Brandeis, concurring in *Whitney v. California*, espoused this view when he stated that "[t]hose who won our independence believed that the final end of the state

---

[118] *See, e.g.*, Eve Thomas, *FREE SPEECH AND ITS RELATION TO SELF-GOVERNMENT. By Alexander Meiklejohn. New York: Harper & Brothers, 1948.*, 3 U. MIAMI L. REV. 66, 66 (1948) (book review).

[119] Mills v. Alabama, 384 U.S. 214, 218 (1966).

[120] *See* Owen M. Fiss, *Free Speech and Social Structure*, 71 IOWA L. REV. 1405 (1986) ("[T]o serve the ultimate purpose of the first amendment we may sometimes find it necessary to 'restrict the speech of some elements of our society in order to enhance the relative voice of others,' and that unless the Court allows, and sometimes even requires, the state to do so, we as a people will never truly be free."); *see also* Robert H. Bork, *Neutral Principles and Some First Amendment Problems*, 47 IND. L. J. 1, 27 (1971) ("Two types of problems may be supposed to arise with respect to this solution. The first is the difficulty of drawing a line between political and non-political speech. The second is that such a line will leave unprotected much speech that is essential to the life of a civilized community.").

was to make men free to develop their faculties, and that in its government the deliberative forces should prevail over the arbitrary."[121]

However, this theory has come under attack for being too broad. According to legal scholar Ronald Bork, the development of individual faculties and the achievement of pleasure through speech is no different from any other human activity.[122] "An individual," according to Bork, "may develop his faculties or derive pleasure from trading on the stock market, following his profession as a river port pilot, working as a barmaid, engaging in sexual activity, playing tennis, rigging prices or in any of thousands of other endeavors."[123] Therefore, speech only serves as the ultimate source of autonomy and self-fulfillment inasmuch as the individual prioritizes it in her own hierarchy of desires.[124]

D. HOW DOES FIRST AMENDMENT THEORY APPLY TO HATE IN CYBERSPACE?

The relationship between First Amendment theory and hate speech is complicated by the fact that—even offline—justifying the presence of hate speech is an arduous task.

For starters, the autonomy justification feels wicked even to argue. The idea that somebody would feel a sense of self-fulfillment from spreading hate speech—whether it be uttering racial epithets, making derogatory statements toward a particular religious group, or burning a cross in somebody's front yard—should not be tolerated in a civilized society.

The democratic self-government justification for protecting speech, on the other hand, does not really apply at all to hate speech used offline. Because much of the case law addressing hate speech involves matters of private concern (e.g., a teenager burning a cross in a neighbor's yard), a self-government theorist—such as Meiklejohn—might agree that the argument for protecting hate speech seems to be lacking under this justification.

---

[121] Whitney v. California, 274 U.S. 357, 375 (1927) (Brandeis, J., concurring).
[122] Bork, *supra* note 120, at 25.
[123] *Id.*
[124] *See id.*

The search for truth in the marketplace of ideas, however, is where proponents of protecting hate speech find their most compelling argument. Real-world experience has shown us that, unfortunately, hate speech laws—which license the government to punish speech solely because its message is disfavored, disturbing, or feared—often are applied disproportionately against unpopular, dissenting views.[125] For example, the very laws that were enacted in Skokie, Illinois, to prohibit neo-Nazis from marching through the city—according to the American Civil Liberties Union—could have been used to stop Martin Luther King, Jr.'s march into Cicero, Illinois, in 1968.[126]

And if social media accounts are analogous to public parks, streets, or other real-world public forums, this argument certainly would have merit. However, simply put, they are not similar enough to justify protecting online hate speech under the marketplace of ideas rationale.

Between 1919 and 1927—the years between the cases in which the marketplace of ideas was conceptualized—spreading an antisemitic message to one billion people in a matter of minutes was impossible. Today, that is theoretically possible: By writing a message, copying and pasting its contents, and posting it on the top ten most-followed Twitter pages,[127] a user could spread hate speech to a billion online profiles in less time than it takes to drive to a restaurant across town.

Justice Kennedy, writing for the majority in *Packingham v. North Carolina*, opined that "[t]he forces and directions of the Internet are so new, so protean, and so far reaching that courts must be conscious that what they say today might be obsolete tomorrow."[128] And we must all be conscious of this fact. Between 2015 and 2021, the total number of social media users grew from 2.07 billion to 4.48 billion;[129] the rapidly changing

---

[125] *See* STROSSEN, *supra* note 63, at 7.

[126] *Id.* at 16.

[127] *See Top Twitter accounts in the world in 2023*, TWEET BINDER, https://www.tweetbinder.com/blog/top-twitter-accounts/ (last visited Jan. 19, 2023).

[128] Packingham v. North Carolina, 137 S. Ct. 1730, 1736 (2017).

[129] *See Social Media Statistics Details*, THE UNIV. OF ME., https://umaine.edu/undiscoveredmaine/small-business/resources/marketing-for-small-business/social-media-tools/social-media-statistics-details/ (last visited Jan. 19, 2023).

information environment has made it easier for misinformation to spread;[130] and a little under half of American adults "often" or "sometimes" turn to social media for their news.[131]

These three trends, when considered together, shed light on a frightening reality: Americans are using social media—which is flooded with misinformed content—to consume information at a rapidly increasing rate. Twenty-six years have passed since *Reno v. ACLU*, in which the Court found that the internet is "not as 'invasive' as radio or television" and is therefore subject to less regulation.[132]

Technology is ever-changing, and the *Reno* Court likely could not have imagined that cell phones—sending notifications to entice you to consume more content—would be owned by a substantial majority of the nation with access to the internet.[133] The Court surely did not foresee the fact that an individual would be able to deliver hate speech, misinformation, etc., into one's home with the click of a button. Nonetheless, this is where the nation stands, and it is still practicable for the conflicting interests of First Amendment theory and online speech to reconcile.

V. A NECESSARY ELEMENT TO REGULATING HATE SPEECH: HOLDING BIG TECH ACCOUNTABLE FOR ITS CONTENT MODERATION SCHEMES

Section 230 of the Communications Decency Act ("CDA") has been welcomed by proponents of protecting online hate speech because it "fosters less speech-restrictive alternatives to government censorship."[134] Under section 230 of the CDA, platforms may allow all constitutionally protected speech, or they may choose to permit only speech that does not

---

130 *See* Emily Bazelon, *The Disinformation Dilemma*, *in* SOCIAL MEDIA, FREEDOM OF SPEECH AND THE FUTURE OF OUR DEMOCRACY 41, 41 (2022).

131 *See* Mason Walker & Katerina Eva Matsa, *News Consumption Across Social Media in 2021*, PEW RSCH. CTR. (Sept. 20, 2021), https://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021/.

132 *See* Reno v. ACLU, 521 U.S. 844, 868–69 (1997).

133 *See Mobile Fact Sheet*, PEW RSCH. CTR. (Apr 7. 2021),https://www.pewresearch.org/internet/fact-sheet/mobile/.

134 *See* Strossen, *supra* note 27, at 8.

contain various controversial topics.[135] Specifically, section 230(c)(2) precludes government regulation of platform content moderation policies.[136]

Some platforms—such as "Rumble" and "Gettr"—identify themselves as havens for free speech, reject "cancel culture," and aim to promote the marketplace of ideas.[137] On the other hand, other platforms—most notably, Facebook—are committed to limiting hate speech on their platforms.[138] And leaving it that way may be best for democracy.

In the debate surrounding content moderation, there are few points of agreement, but a common thread across the board is a distrust of how companies arrive at their content moderation decisions.[139] Contemporary discussion of content moderation is misleading and incomplete.[140] Regulators and scholars alike assume that content moderation is a set of legislative-style substantive rules applied to individual cases, where ex-post review effectively solves disputes.[141] However, Professor Evelyn Douek set out a framework for moderating content that focuses on the content moderation picture before cases arrive, which is where the "most important decisions in content moderations happen. . . ."[142]

Section 230(b) of the CDA, among other things, declares that it is the policy of the United States to promote the continued development of the internet and to preserve the vibrant, competitive free market that "presently exists" for the

---

135 *See id.*

136 Ceresney et al., *supra* note 2, at xxvii.

137 *See* Galen Stocking et. al, *Alternative social media sites frequently identify as free speech advocates*, PEW RSCH. CTR. (Oct. 6, 2022), https://www.pewresearch.org/journalism/2022/10/06/alternative-social-media-sites-frequently-identify-as-free-speech-advocates/.

138 *See Hate Speech*, META, https://transparency.fb.com/policies/community-standards/hate-speech/ (last visited Feb. 5, 2023).

139 *See* Evelyn Douek, *Content Moderation as Systems Thinking*, 136 HARV. L. REV. 528 (2022).

140 *See id.* Content moderation may be defined as the "governance mechanisms that structure participation in a community to facilitate cooperation and prevent abuse." *See* James Grimmelmann, *The Virtues of Moderation*, 17 YALE J. L. & TECH. 42, 47 (2015).

141 *See* Douek, *supra* note 139, at 529.

142 *See id.* at 530.

internet.[143] Unfortunately, the current state of social media content moderation likely inhibits these goals.

Currently, platforms must balance the interests of those opposed to hate speech and those who want unfettered speech online. It is abundantly clear that neither side of the debate is satisfied at the moment.[144] Furthermore, the sensitive nature of the debate can be seen in the mass exodus of Twitter users once Elon Musk began making arbitrary content moderation policy decisions.[145] As a result, dissatisfaction with platforms hinders continued development and use of the internet; simultaneously, many platforms are not allowing a free market of ideas. Thus, the policy goals intended by the CDA are in a state of disarray.

Professor Douek's framework may provide a means for ensuring all parties are satisfied. At the root of her plan is accountability.[146] The following summary of Professor Douek's content moderation proposal contains components that regulate hate speech while preventing institutional bias and achieving legitimacy in the eyes of the public—striking a balance that is essential to the future of content moderation.

First, and most importantly, it is a way to lessen the role that bias plays in the content moderation process. Structural separation, according to Douek, would help eliminate the incentives that make biased conduct possible in the first place.[147] Maintaining separate teams within a company may be a costly task, but a content moderation department separate from potentially biased individuals gives Big Tech companies a sense of procedural fairness unknown to the current moderation processes.[148] Furthermore, by achieving structural separation, institutions will be a step closer to solving a core problem for many companies: simultaneously being "responsible for the rules of the platform and keeping governments happy."[149]

---

[143] *See* 47 U.S.C.A. § 230(b)(1)–(2).
[144] *See* Douek, *supra* note 139, at 528.
[145] *See* Jelani Cobb, *Why I Quit Elon Musk's Twitter*, NEW YORKER (Nov. 27, 2022), https://www.newyorker.com/news/daily-comment/why-i-quit-elon-musks-twitter.
[146] *See* Douek, *supra* note 139, at 586.
[147] *See id.*
[148] *See id.* at 587.
[149] *See id* at 588–89 (quoting former Facebook Chief Security Officer Alex Stamos).

In addition to maintaining structural separation of a company's departments, having clear, transparent procedures is vital to the success of any content moderation plan. Instead of using ex-post review, lawmakers should require platforms to engage in ex-ante planning and risk assessment.[150] This would require platforms to publicly explain their rules, how they will enforce them, and how they will guard against risks to such enforcement.[151] And by requiring platforms to be able to demonstrate they have a system in place to enforce their public commitments with consistency and accuracy,[152] social media companies could drastically improve their perception in the eyes of the public.

Furthermore, implementing this procedure would make a difference in the platforms' internal operations as well. For example, during the 2020 election, Facebook invested significantly in its preparation for information incidents, and observers noted evidence of their improvement in enforcement of rules.[153] YouTube, on the other hand, engaged in a far more perfunctory plan, and the errors that Facebook avoided seemed to fall right into YouTube's hands.[154]

Another consideration is the public's desire to have a say in content moderation. Following Facebook's ban on President Trump, the platform received over 9,000 comments from interested parties; considering the heat of the current political climate, we can expect decisions and responses like this again.[155] Thus, having an ongoing planning and review process would allow platforms to be more consistent, transparent, and prospective.[156]

Lastly, social media companies should establish a form of regulatory system. The closest parallel that leverages private self-regulation and transforms it into a public regulatory framework is in the realm of privacy regulation.[157] The Federal Trade Commission ("FTC") began enforcing companies' self-regulatory policies, essentially providing enough oversight that people would view the companies' policies as legitimate and

---

150 *See id* at 593.
151 *See id.*
152 *See id* at 595.
153 *See id* at 596.
154 *See id.*
155 *See id.*
156 *See id.* at 598.
157 *See id.* at 603.

trustworthy.[158] Similarly, content moderation policies—whether overseen by the FTC or another regulatory agency—would be given extra legitimacy if they had some form of accountability behind them.[159] Interestingly, this would require social media platforms to agree to conform to a new regulatory regime. However, for the platforms who are committed to moderating content—and not permitting all forms of speech—this would be another way of making its products attractive to users who want a hate speech-free platform.

VI. CONCLUSION

To this day, First Amendment jurisprudence protects the rights of speakers whose views may be distasteful or discomforting to the overwhelming majority of people—at least offline. The internet and social media, however, present a complex case for whether or not to regulate hate speech online.

Social media has done a tremendous deal of good in the world by serving as the foundation for promoting awareness of police brutality and the #MeToo Movement. Nevertheless, social media has seen a tsunami of hate since the beginning of the COVID-19 pandemic, and the situation is only escalating. Worst of all, the social media platforms regulating speech online often make arbitrary and uninformed content moderation decisions, resulting in these platforms censoring anti-racism activists for speaking out against white supremacy.

In light of this issue, however, we need to reject calls to prohibit private platforms from moderating content. In the event social media platforms are subject to First Amendment principles, hate speech victims will not receive adequate protection from threatening comments.

And although we should not give in to the calls to permit uninhibited speech on social media, we are not without a solution. We can, in fact, censor the "thought that we hate" on social media, and there is a framework with which to do so—promoting transparency, procedural rights, and oversight, as well as limiting institutional bias in making content moderation decisions. Now it is time for Americans to act and demand respect and responsibility on social media.

---

158 *See id.*

159 *See id.* at 605.