"ALTERNATIVE FACTS" AND HATE: REGULATING CONSPIRACY THEORIES THAT TAKE THE FORM OF HATEFUL FALSITY

SAMANTHA HAY

I. INTRODUCTION

Prepare for the invasion! Hurry, a caravan of migrants, MS-13 gang members, and terrorists forcefully march on our southern border. Don't be fooled by their asylee disguise and the children that accompany them; these people are dangerous and well funded by Jewish globalist George Soros. They are coming to invade our country.

This is the all too familiar caravan narrative. It is a falsity, built upon the large group of migrants journeying to the United States in the hopes of declaring asylum. Both the invasion itself and the Jewish globalists' role as puppeteers are falsehoods. The "narrative" is also an expression of hate—hate of immigrants, Central and Latin Americans, and Jews. The false narrative spread from conspiracy theorists' inner circles, making its way to widely followed media outlets, cable news, and then President Trump's Twitter feed.[1]

USA Today traced the conspiracy theory's focus on George Soros' involvement to an internet post from October 14, 2018, by a conspiracy theorist with six thousand followers who frequently posts about "white genocide."[2] Within hours, it appeared on six Facebook pages whose membership totaled over 165,000 users; just four days later this portion of the conspiracy theory reached over two million people.[3] It continued to spread like wildfire from there, making its way to mainstream platforms. This speech is false speech. This speech is hateful speech. Moreover, this speech harms.

---

[1] Jeremy W. Peters, *How Trump-Fed Conspiracy Theories About Migrant Caravan Intersect With Deadly Hatred*, N.Y. TIMES (Oct. 29, 2018), https://www.nytimes.com/2018/10/29/us/politics/caravan-trump-shooting-elections.html.

[2] Brad Heath et al., *How a Lie About George Soros and the Migrant Caravan Multiplied Online*, USA TODAY (Oct. 31, 2018, 12:27 PM), https://www.usatoday.com/in-depth/news/nation/2018/10/31/george-soros-and-migrant-caravan-how-lie-multiplied-online/1824633002/. White genocide is a popular white supremacy conspiracy theory alleging that the "white race is 'dying' due to non-white populations 'forced assimilations' all of which are deliberately engineered and controlled by [Jews]." *White Genocide*, ANTI-DEFAMATION LEAGUE, https://www.adl.org/resources/glossary-terms/white-genocide (last visited Dec. 13, 2018).

[3] Heath et al., *supra* note 2.

The harm that "hateful falsity" creates is antisocial behavior[4]—behavior at odds with societal norms—through its cognitive consumption process.[5] Antisocial behavior, as an epidemiological phenomenon, endangers society.[6] Hateful falsity causes antisocial dispositions through its cognitive consumption process, the hubs in which hateful falsity is disseminated (internet echo chambers), and its self-sealing irreparable qualities.[7] Individuals exhibiting antisocial behaviors are more paranoid, aggressive, and violent than society as a whole.[8] To ensure safe, prosperous, and communal living, government has a compelling interest in addressing—head-on—antisocial behavior and its causes.

I propose that government take this antisocial behavior on by regulating hateful falsity. This proposal rests on understanding how hateful falsity causes antisocial behavior and the dangers of its growth. It also requires asking whether the First Amendment allows government to step in and act. Hateful falsity is, by definition, false speech, a factor that supports the constitutionality of its regulation.[9] Falsity alone, however, is not enough in light of the Supreme Court's established hostility to regulations on hate speech[10] and its regular and fundamental application of content-based analysis, also known as the "cardinal rule," to regulations distinguishing speech on the basis of its content.[11]

Rather than dissect the various First Amendment doctrines implicated by hateful falsity (e.g., false speech, hate speech, speech on the internet, and defamation), I encourage a more comprehensive change to the framework. I argue that because regulating hateful falsity has a non-censorial purpose[12]—meaning it does not aim to suppress the speech because of its message but rather seeks to combat the harm of growing antisocial behavior—it warrants government regulation. Despite the Supreme Court's indiscriminate and careless equivalenting between content-based regulations that suppress speech because of its message and suppression for the purpose of addressing real harm, a content-based regulation with a non-censorial purpose can survive judicial review because this regulation addresses the real social harm of antisocial behaviors.

I turn to the political theories and values of free speech protection and argue that unregulated hateful falsity undermines First Amendment values. Some of these theories support such regulation because of an inability to

---

[4] *See* Sander van der Linden, *The Conspiracy-Effect: Exposure to Conspiracy Theories (about global warming) Decreases Pro-Social Behavior and Science Acceptance*, 87 PERS. & INDIVIDUAL DIFFERENCES 171, 172 (2015); Jan-Willem van Prooijen & Karen M. Douglas, *Belief in Conspiracy Theories: Basic Principles of an Emerging Research Domain*, 48 EUR. J. SOC. PSYCHOL. 897, 902–04 (2018) (noting that affirmatively defining a causal relationship requires additional research).

[5] *See infra* Part III.A.

[6] *See* S. Alexandra Burt & Jenae M. Neiderhiser, *Aggressive Versus Nonaggressive Antisocial Behavior: Distinctive Etiological Moderation by Age*, 45 DEVELOPMENTAL PSYCHOL. 1164, 1–2 (2009).

[7] *See infra* Part III.A.

[8] *See* van der Linden, *supra* note 4, at 172.

[9] The Supreme Court's treatment of false speech highlights this. *See infra* Part IV.B(i).

[10] The Court is hostile to hate speech regulations because it is suspicious that the regulations' content-based distinctions seek to restrict ideas. *See generally* R.A.V. v. City of St. Paul, 505 U.S. 377 (1992) (striking down as an impermissible content-based regulation a prohibition on the placement of hateful symbols on another's property).

[11] Rebecca L. Brown, *The Harm Principle and Free Speech*, 89 S. CAL. L. REV. 953, 954–55 (2016) (referring to content-based analysis as the "'cardinal rule' of free speech" jurisprudence).

[12] *Id*. at 960–61, 961 n.35.

further First Amendment values in any other way. Alternative theories demonstrate why the "more speech not less" model fails when it comes to hateful falsity. The Court has recognized government's interest in curbing social harms as compelling. Moreover, a proper regulation can be narrowly tailored to address antisocial behavior without chilling or burdening constitutionally protected speech.

II. WHAT IS HATEFUL FALSITY?

As apparent by its name, hateful falsity encompasses speech that is both hate speech and false speech. I define hateful falsity as demonstrably and materially false conspiracy theories that incorporate expressions of hate on the basis of race, religion, ethnicity, sexual orientation, sex, and gender identity. Hateful falsity distinguishes between statements of fact and expressions of ideas and opinions. Because the human form of expression defies strict categorization, hateful falsity refers only to statements that purport to offer a true and factual rendition of events in contrast to expressions of opinion that falsely characterize facts. How the speech is presented, delivered, and disseminated informs this distinction as well. Is the speech presented as a piece of news and a rendition of truth, or is it an expression of one's thoughts, ideas, and criticism? Moreover, a certain degree of falsity, such as materiality, is necessary to qualify as hateful falsity, thereby ruling out inadvertent misstatements.

To highlight the breadth of hateful falsity's harm, I offer examples where hateful falsity has led to violence and examples where it has not. The migrant caravan example is one in which hateful falsity directly contributed to violence—a mass shooting in a Jewish synagogue, an attempted bombing of a Jewish philanthropist, and likely acts of violence on Mexicans and Central Americans.[13]

I offer another example of hateful falsity where, violence did not directly result, yet hateful falsity still caused harm. A Facebook meme of Congresswoman Ilhan Omar, a Somali-American Muslim refugee who wears a head covering, quoted her as saying, "I am America's hope and the [P]resident's nightmare . . . I think all white men should be put in chains as slaves because they will never submit to Islam."[14] Context makes clear that the meme is not a parody. Omar did say, "I am America's hope and the president's nightmare" during an interview on *The Daily Show with Trevor Noah*.[15] Omar, however, has never called for enslaving white men or punishing those who do not practice Islam.[16] PolitiFact verified that no such quote exists.[17]

No violence was perpetrated in *direct* response to this grotesque and false meme. Yet, even in the absence of directly attributable violence, the

---

[13] Peters, *supra* note 1.

[14] Miriam Valverde, *Facebook Meme Falsely Attributes Quote About Race, Slaves and Islam to Ilhan Omar*, POLITIFACT (Nov. 15, 2018, 4:01 PM), https://www.politifact.com/facebook-fact-checks/statements/2018/nov/15/viral-image/facebook-meme-falsely-attributes-quote-ilhan-omar/.

[15] *Id*.

[16] *Id*.

[17] PolitiFact found no evidence of the quote and designates the statement "pants on fire," the highest tier of falsity. *Id*.

meme's hateful falsity is still extremely harmful because of its antisocial effects. The meme falsely claims that a Muslim woman, participating in representative democracy, supports white male enslavement and violent adherence to Islam. Not only does this perpetuate Islamophobia and sexism as well as bolster other conspiracy theories such as white genocide, this hateful falsity further erodes the audience members' ability to engage in rational analytical processes and deduce untruths. Each building block nurtures the development of antisocial behavior.

Below, I explain how exposure to hateful falsity causes antisocial behavior. I pause now to highlight the sufficient but not necessary role of violence in hateful falsity. I emphasize that all hateful falsity causes antisocial behavior even if only some are also *directly* linked to violence. One is not driven to mass murder from an isolated exposure to hateful falsity. Rather, each hateful falsity increases antisocial behavior and tendencies, and sometimes that results in unspeakable violence. The harm perpetrated by the migrant caravan conspiracy theory and Omar's meme is the development of dangerous patterns of antisocial behavior. This harm justifies a government response.

III. THE HARM: HATEFUL FALSITY CAUSES ANTISOCIAL BEHAVIOR

Hateful falsity certainly hardens our divides, nurtures intolerances, and strengthens our hate. It offends, humiliates, and belittles its victims; these sorts of harms stem from abhorrent ideas. They, however, are not the harms I seek to address with regulation. Hateful falsity causes another type of harm: a social harm. At its core, it breeds dangerous antisocial behavior. I imagine a decent portion of hateful falsity exists to incite discrimination and racial hostility. The goal of the speaker is, however, irrelevant to me. Rather, I am interested in its effects—antisocial behavior. I propose regulating hateful falsity not because of its abhorrent ideas; I do not seek to counter the speaker's efforts to persuade people of their ideas. Whether hateful falsity is disseminated for ideological or financial gain is irrelevant; its harm is the target. I argue for regulation because I believe the government has a compelling interest in protecting against the harms of antisocial behaviors caused by hateful falsity.

A. THE PSYCHOLOGY OF HATEFUL FALSITY

Psychologists, researchers, and academic scholars have recently increased their focus on conspiracy theories.[18] Not all conspiracy theories constitute hateful falsity; nevertheless, research on conspiracy theories helps explain how hateful falsity causes antisocial behavior.[19] Given that conspiracy theories are especially susceptible to hateful expression, the

[18] *See* William Cummings, *Conspiracy Theories: Here's What Drives People to Them, No Matter How Wacky*, USA TODAY (Dec. 23, 2017), https://www.usatoday.com/story/news/nation/2017/12/23/conspiracy-theory-psychology/815121001/.

[19] Most conspiracy theories do not fall into the definition of hateful falsity. *Id*. Not all conspiracy theories are expressions of hate. *Id*. Moreover, some theories that once seemed conspiratorial have turned out to be true, such as the CIA's testing of LSD on unknowing individuals. *Id*.

research on all conspiracy theories is particularly applicable to hateful falsity.[20] I explore how individuals come to believe conspiracy theories, how conspiracy theories spread, and the increased rate at which people come to believe conspiracy theories as they consume more. Alongside this information, I address the implications of the exposure to hate speech to highlight how the combination of the two is especially harmful.

Psychologist Sander van der Linden defines conspiracy theories as theories that suggest "[s]ome covert and powerful individual(s), organization(s), or group(s) are intentionally plotting to accomplish some sinister goal."[21] There are many iterations of the definition, though all identify a belief that a group of powerful people is orchestrating a secret and destructive plot.[22]

Psychologists explain that "the human brain is wired to find conspiracy theories appealing."[23] Susceptibility to conspiracy theories is not limited to individuals suffering from paranoia, delusions, or other mental illness; individuals who do not suffer from mental illness also believe conspiracy theories devoid of supporting evidence, even when ample evidence exists to dispel the theory.[24] In fact, more than half of Americans believe in at least one conspiracy theory.[25] No research, however, exists that captures what percentage of Americans believe in a hateful falsity specifically.

Individuals are susceptible to give in and believe conspiracy theories because of a combination of "cognitive bias[es]": "confirmation bias," the tendency to embrace explanations consistent with the individual's established beliefs; "proportionality bias," the tendency to "believe big events must have big causes"; and "illusory pattern perception," the inclination to suspect a causal relationship where one does not exist.[26] The presence of these biases increases acceptance of the conspiracy theory.[27] Alongside these biases, individuals who heavily rely on gut feelings and instinct over rational deliberation and analytical thinking are more likely to succumb to conspiracy theories. As individuals accept conspiracy theories as true, they adopt a more conspiratorial view of the world eroding their ability to engage in rational deliberation in sacrifice of self-serving beliefs consistent with their larger conspiratorial perspective.[28]

Conspiracy theories are first accepted by individuals "with low thresholds for acceptance."[29] As those with low acceptance thresholds adopt the theory, "informational pressure builds," and those with higher thresholds of acceptance are more likely to be persuaded by seeing others support the

---

[20] *See* Viren Swami, *Social Psychological Origins of Conspiracy Theories: The Case of the Jewish Conspiracy Theory in Malaysia*, 3 FRONTIER PSYCHOL. 1, 1 (2012).

[21] Van der Linden, *supra* note 4, at 171.

[22] *See, e.g.*, Cass R. Sunstein & Adrian Vermeule, *Conspiracy Theories: Causes and Cures*, 17 J. POL. PHIL. 202, 207 (2009) ("conspiracy theories generally attribute extraordinary powers to certain agents—to plan, to control others, to maintain secrets, and so forth"); Swami, *supra* note 20, at 1.

[23] Cummings, *supra* note 18.

[24] *See* Sunstein & Vermeule, *supra* note 22, at 211–12.

[25] Van der Linden, *supra* note 4, at 171.

[26] Cummings, *supra* note 18; Christopher French, *Why Do Some People Believe in Conspiracy Theories?*, SCI. AMERICAN (July 1, 2015), https://www.scientificamerican.com/article/why-do-some-people-believe-in-conspiracy-theories/.

[27] Cummings, *supra* note 18.

[28] *See id.*; *see also* van der Linden, *supra* note 4, at 171.

[29] Sunstein & Vermeule, *supra* note 22, at 214.

theory.[30] Influenced by an interest in preserving a non-naïve reputation and group polarization, more people are led to believe the theory. Evidence demonstrates that individuals with polarizing beliefs, a trend as of late, are more likely to believe a conspiracy theory after speaking with individuals with those same polarizing beliefs.[31] As the polarization rises, these groups become more committed and self-segregating, either physically or in access to information.[32] In sum, the number of people exposed to a conspiracy theory impacts an individual's ability to resist the conspiracy theory.

The more an individual believes in a conspiracy theory, the more susceptible he or she is to believing new conspiracy theories.[33] Put simply, conspiracy theories breed more conspiracy theories.[34] When two people are exposed to a conspiracy theory, the individual who believes in ten conspiracy theories is more likely to believe the new theory than the individual who believes in just one. Van der Linden characterizes this phenomenon as a "slippery slope"[35] and warns that conspiracy theories "spread quickly and can do more harm than you think."[36] In other words, conspiracies can have a compounding effect on individuals and the public. Therefore, effectively addressing the ultimate harm of hateful falsity requires stepping in at its earliest public exposure.

Given these compounding effects and the negative feedback loop, disproving conspiracy theories is an uphill challenge. Conspiracy theories are not put to rest when presented with clear and numerous facts, direct denials, or counter speech.[37] Strong believers respond to counter-speech and counter-evidence by attempting to discredit the source or by labeling such evidence a so-called cover-up.[38] Take, for example, conspiracies surrounding climate change and President Obama's birthplace.[39] The tendency to cling to conspiracy theories has been described as a "sort of psycho-religious belief."[40] Therefore, to address the harm of hateful falsity and its resulting antisocial behavior, government must regulate hateful falsity at its early public exposure.

B. HATEFUL FALSITY CAUSES ANTISOCIAL BEHAVIOR

Antisocial behavior is the tendency to engage in acts at odds with social norms, sometimes in aggressive, interpersonally invasive, and violent ways.[41] Antisocial behavior often takes the form of aggression, violence, bullying, manipulation, and law or rule-breaking.[42] Antisocial behavior

---

[30] *Id*.
[31] *See id*. at 212–14, 216–17.
[32] *Id*. at 217–18.
[33] Van der Linden, *supra* note 4, at 171–73.
[34] *Id*.
[35] *Id* at 171.
[36] Sander van der Linden, *The Surprising Power of Conspiracy Theories*, PSYCHOL. TODAY (Aug. 24, 2015), https://www.psychologytoday.com/us/blog/socially-relevant/201508/the-surprising-power-conspiracy-theories.
[37] *See id*.
[38] *See* Sunstein & Vermeule, *supra* note 22, at 210, 221–23.
[39] *See* Cummings, *supra* note 18.
[40] *Id*.
[41] *See* Burt & Neiderhiser, *supra* note 6, at 1164.
[42] *Antisocial Behavior*, PSYCHOL., http://psychology.iresearchnet.com/social-psychology/antisocial-behavior/ (last visited Dec. 13, 2018).

manifests in both overt and covert forms. Antisocial Personality Disorder is a designated disorder in the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders ("DSM-IV")*.[43] While the role of nature versus nurture is still debated by psychologists, the evidence makes clear that certain environments can trigger antisocial behavior.[44]

The link between conspiracy theories and antisocial behavior is far from theoretical. Exposure to conspiracy theories subconsciously alters individuals' attitudes.[45] Evidence shows that such exposure makes individuals "less pro-social and less willing to contribute to important societal causes."[46] As the intake of conspiracy theories grow, we see paranoid cogitation and increased distrust in institutional and societal structures (i.e., antisocial behavior).[47] The consequences of antisocial behavior continue to grow from there.

Sometimes, the antisocial behavior that conspiracy theories incite can result in violence. For example, the individuals who committed the Oklahoma City bombings shared beliefs in government conspiracy theories.[48] In October 2018, the man who targeted the Jewish synagogue on Shabbat (the Jewish Sabbath) shouted, "All Jews must die," before murdering eleven individuals.[49] Moments before his violence, the shooter expressed his motivation on social media.[50] He believed the hateful falsity that Jews were orchestrating a mass migration scheme to bring non-white immigrants to the United States to the detriment of white Americans.[51] Here, hateful falsity was a contributory factor in causing antisocial behavior that resulted in mass murder.

Hateful falsity causes antisocial behaviors that result in other types of violence as well. Hate crimes reported in 2017 increased by 17 percent. Within this figure was a 63 percent increase in reported hate crimes against American Indians or Alaska Natives, a 37 percent increase in reported hate crimes against Jews, and a 24 percent increase in hate crimes against Hispanic and Latino individuals.[52] While it is unclear how many of these crimes were motivated in part to hateful falsity, the rise of hateful falsity suggests a relationship.

The additional dimension of hate to the falsity makes the antisocial behavior all the more dangerous. On its own, the research shows how

---

[43] *Id*.

[44] *Id*.

[45] *See* David Jolley & Karen M. Douglas, *The Social Consequences of Conspiracism: Exposure to Conspiracy Theories Decreases Intentions to Engage in Politics and to Reduce One's Carbon Footprint*, 105 BRITISH J. PSYCHOL. 35, 37 (2014).

[46] Van der Linden, *supra* note 4, at 172. Conspiracy theories have also been defined as "a subset of false narrative in which the ultimate cause of an event is believed to be due to a malevolent plot by multiple actors working together." Swami, *supra* note 20, at 1.

[47] Swami, *supra* note 20, at 13.

[48] Sunstein & Vermeule, *supra* note 22, at 220.

[49] David Lind, *The Conspiracy Theory that Led to the Pittsburgh Synagogue Shooting, Explained*, VOX (Oct. 29, 2018, 3:20 PM), https://www.vox.com/2018/10/29/18037580/pittsburgh-shooter-anti-semitism-racist-jewish-caravan.

[50] *Id*.

[51] *Id*.

[52] German Lopez, *FBI: Reported Hate Crimes Increased by 17 Percent in 2017*, VOX (Nov. 13, 2018, 1:10 PM), https://www.vox.com/policy-and-politics/2018/11/13/18091646/fbi-hate-crimes-2017. The FBI published these statistics and noted that the increase may in part be due to an increase in reporting rather than an increase in actual hate crimes. *Id*.

dangerous general conspiracy theories are to society as a whole. Adding expression of hate only aggregates the danger. From a First Amendment perspective, hate speech is not precisely defined. Generally, hate speech expresses hatred and prejudice towards an individual or group of individuals based on race, religion, ethnicity, sexual orientation, sex, or gender identity.[53] Often, hate speech is communicated through epithets or symbols but can also take the form of historical revisionism or scientific data.[54]

Frequent exposure to hate speech desensitizes its audience. Initially, audience members' "affective [sic] responses to . . . hostile messages" are reduced until they eventually reach "a tipping point," and "hate speech. . . [is] interpreted . . . as less negative and harmful, less important, and less violating of social norms."[55] Professor Wiktor Soral explains that exposure to hate speech desensitizes the audience members to its offensiveness, decreases sympathy for hate speech victims, and increases their prejudices.[56] The paranoia, anger, and aggression of the antisocial disposition now have specific targets and scapegoats.

IV. WHAT CAN GOVERNMENT DO? REGULATE IT

The most effective way to address antisocial behavior and minimize the concomitant social harm is to regulate hateful falsity. Combatting the antisocial behavior caused by hateful falsity requires addressing it at its earliest public dissemination. Regulation, therefore, should impose sanctions on the public dissemination of hateful falsity. Nevertheless, the First Amendment may hinder government's ability to regulate hateful falsity. I explain why a non-censorial regulation of hateful falsity seeking to address real social harm survives constitutional muster. To facilitate a clear analysis of the constitutional issues and arguments in favor of regulation, I sketch out what a regulation of hateful falsity might look like.

A. SAMPLE REGULATION

The purpose of regulating hateful falsity is not to censor or suppress unfavorable ideas. The fact that hateful falsity is ugly and that its *words* are antithetical to many constitutional values are not reasons to regulate hateful falsity. Rather, antisocial behavior—the effect of hateful falsity—is the reason hateful falsity should be regulated. The regulation must, therefore, reflect this purpose and serve this aim.

To protect ideas and avoid government censorship, a regulation of hateful falsity must be well-conceived, laser-focused, and clear. I offer some possible features of a regulation. When discussing the constitutional questions, I refer to these features and how they prevent overbreadth, the chilling of protected speech, and viewpoint discrimination. Given the

---

[53] Laws regulating hate speech have been defined as "[l]aws that prohibit the expression of hate, commonly called hate speech, against individuals or groups based on national or ethnic origin, race, or religion . . . ." John C. Knechtle, *When to Regulate Hate Speech*, 110 PENN ST. L. REV. 539, 539 (2006).

[54] *See id*.

[55] Wiktor Soral et al., *Exposure to Hate Speech Increases Prejudice Through Desensitization*, 44 AGGRESSIVE BEHAV. 136, 137 (2018).

[56] *Id*. at 141.

fundamental role of false speech in hateful falsity, I borrow many of the standards central in U.S. libel laws.

The regulation of hateful falsity should take the form of a prohibition enforced through sanctions. The regulation should impose liability on those who publicly create or disseminate demonstrably and materially false statements purporting to be true renditions of events and that incorporate expressions of hate towards a race, ethnicity, religion, sex, gender identity, or sexual orientation. Moreover, the regulation should have a *mens rea* requirement such as knowledge. To be subject to sanctions, the speaker must *know* that the statement is false and choose to disseminate it recklessly anyways.

Moreover, requiring that the speech purport to be a true statement of fact and thereby excluding normative ideas or opinions by definition excludes things like satire, art, literature, and research. By requiring that the statement was demonstrably false, its falsity deemed material to the message, and the knowledge requirement satisfied, prevents inadvertent misstatements, hyperbole, sloppy research, and poor fact-checking from falling victim to the regulation.

Government must bear the burden of proof and prove each of these elements by clear and convincing evidence for civil liability or hypothetically beyond a reasonable doubt for criminal liability.[57] To be sure, this would require juries to engage in factfinding, including whether a statement is a fact or opinion and whether it is true or demonstrably and materially false. Juries are certainly accustomed to this type of factfinding.[58] Another possible feature that would offer additional protection is the implementation of a mandatory *de novo* appellate review following a finding of hateful falsity.[59]

Given that public dissemination is integral to the harm of antisocial behavior, the regulation ought to address speech publicly disseminated through the internet, broadcast, and radio as opposed to speech made outside these mediums. As with every regulation of speech, each statutory term and element should be well-defined to avoid vagueness challenges. Establishing liability for hateful falsity ought to be difficult for government, while simultaneously allowing government to protect against hateful falsity's harm.

The regulation might read: *no person shall knowingly or intentionally create for public dissemination or publicly disseminate demonstrably and materially false statements purporting to be true that incorporate expressions of hate on the basis of race, ethnicity, religion, sex, gender identity, or sexual orientation. Satirical expressions, hyperbole, science, art, literature, and research shall be excluded from this prohibition.* As stated above, this is a sample regulation offering possible features of a regulation for the analysis below. The paramount necessity of targeting the speech most

[57] This Note does not take a position as to whether the regulation should be limited to civil liability. Further research by policy analysts is necessary to determine which sanction is necessary to achieve its goal. Surely, a criminal penalty would receive more severe scrutiny from the Supreme Court.

[58] One possible standard is the standard applied to commercial warning labels.

[59] *See* Bose Corp. v. Consumers Union of U.S., Inc., 466 U.S. 485, 514 (1984) (holding a *de novo* review is required for a finding of actual malice).

capable of causing the concerning antisocial behaviors requires policymakers to draft a statute that most effectively combats such harms.

B. OVERCOMING FIRST AMENDMENT CHALLENGES

Any attempt to regulate hateful falsity would be vulnerable to First Amendment attack. But such vulnerability does not render all such regulations fatal or futile. As discussed further below, there are reasons to believe that such regulation could survive a constitutional challenge, including: the Supreme Court's treatment of false speech, the non-censorial nature of the proposed regulation, the regulation's adherence to the theoretical justifications for protecting free speech, and the severe social consequences of antisocial behavior and government's compelling interest in combatting that harm.

1. Falsity

Despite the increased attention to "fake news," false speech is not a new constitutional issue.[60] Dean Erwin Chemerinsky points out that "Fake News" is false speech and dates back to our nation's earliest years when Congress passed the 1789 Alien and Sedition Acts that, in part, prohibited false writing against the government. Under the Alien and Sedition Acts, Congress:

> prohibited the publication of 'false, scandalous and malicious writing or writings against the government of the United States, or either house of the Congress of the United States, or the President of the United States, with intent to defame . . . or to bring them . . . into contempt or disrepute; or to excite against them . . . hatred of the good people of the United States, or to stir up sedition within the United States, or to excite any unlawful combinations therein, for opposing or resisting any law of the United States, or any act of the President of the United States.[61]

It is noteworthy that the Congress that enacted these laws included many of the Framers who drafted, supported, and ratified the Constitution.[62] The Supreme Court never considered the constitutionality of the Alien and Sedition Acts before their repeal.

False speech's First Amendment status, however, remains somewhat unclear[63] as the Court's jurisprudence on false speech tends to fluctuate. The Court refuses to categorically exclude false speech from First Amendment protection.[64] Justice Brennan famously declared that an "erroneous statement is inevitable in free debate, and that it must be protected if the freedoms of expression are to have the 'breathing space' that they need to survive."[65] We tolerate some falsity on the periphery because we are unwilling to sacrifice

---

[60] Erwin Chemerinsky, *False Speech and the First Amendment*, 71 OKLA. L. REV. 1, 1 (2018).
[61] *Id*. (citing Ch. 74, 1 Stat. 596 (1798)).
[62] *Id*.
[63] *Id*. at 2, 5.
[64] United States v. Alvarez, 567 U.S. 709, 718 (2012).
[65] New York Times Co. v. Sullivan, 376 U.S. 254, 271–72 (1964).

any true speech.[66] This breathing space reflects an instinct that some false speech plays an important role in protecting true high-value speech. In other words, without protecting some false speech, public discourse would be hampered because people would be deterred from speaking unless they are certain the speech was true.

The Court also explained that an "erroneous statement of fact is not worthy of constitutional protection," but "[t]he First Amendment requires that we protect *some* falsehood in order to protect speech that matters" because "punishment of error runs the risk of inducing a cautious and restrictive exercise of the constitutionally guaranteed freedoms of speech and press."[67] When the Court rejected *per se* liability for falsity and defamation publications of public figures[68] and introduced the actual malice standard, it invited a certain amount of false speech into the periphery of true speech and into constitutional protection.[69] This prevents the deterrence of high-value speech and inadvertent errors that can too easily enable government censorship of speech at the heart of the First Amendment. In essence, we tolerate some false speech, not because we value falsity, but because it serves as an important buffer.

Nevertheless, the Court's record suggests that false speech that *actually* harms is regulable. There is a longstanding historical view that false and defamatory statements can perpetrate harms,[70] and that falsity by itself is of little value. In these instances, the Court has permitted the regulation of false speech. For example, we sanction individuals who make false defamatory statements against another because of the actual harm to a person's reputation.[71] The Court introduced a more rigorous "actual malice" standard for libel cases brought by public officials in *New York Times v. Sullivan* by requiring public officials to prove, by clear and convincing evidence, (1) the falsity of the statement and (2) actual malice, meaning that the defendant knew the statement was false or acted with reckless disregard for the truth.[72]

In *New York Times*, the Montgomery County Commissioner L.B. Sullivan sued *The New York Times* for libel based on its advertisement that was critical of southern segregationists' "wave of terror" opposition to affording equal rights to black citizens.[73] Sullivan claimed the advertisement defamed him as he was a representative for Montgomery; the advertisement described multiple non-violent protests and the violent responses that followed.[74] Some inaccuracies existed: protestors sang the national anthem,

---

[66] *See* Martin H. Redish & Kyle Voils, *False Commercial Speech and the First Amendment: Understanding the Implications of the Equivalency Principle*, 25 WM. & MARY BILL RTS. J. 765, 767 (2017) ("Even though false speech in and of itself serves no value and often causes harm, occasions will arise in which false speech must be protected in order to foster broader values and societal needs.").

[67] Gertz v. Robert Welch, Inc., 418 U.S. 323, 340–41 (1974) (emphasis added) (holding that actual malice is not required in a libel suit brought by an individual who is not a public official or public figure).

[68] The Court reviewed a finding of libel under an Alabama law which identified "a publication [as] 'libelous per se' if the words 'tend to injure a person in his reputation' or to 'bring [him] into public contempt.'" *New York Times*, 376 U.S. at 267 (second alteration in original) (internal citation omitted).

[69] *See generally id.*

[70] *See* Brown, *supra* note 11, at 984–85.

[71] *See New York Times*, 376 U.S. at 277–83; *see also* Cass R. Sunstein, *Hard Defamation Cases*, 25 WM. & MARY L. REV. 891, 891–92 (1984).

[72] *See generally* New York Times Co. v. Sullivan, 376 U.S. 254 (1964).

[73] *See generally id.*

[74] *See generally id.*

not "My Country Tis of Thee"; Dr. King had been arrested on four occasions rather than seven; and the cause for students' expulsion was a lunch counter-protest, not a demonstration on the Capitol.[75] Based on these inaccuracies, Alabama courts ordered *The New York Times* to pay Sullivan damages.[76] Sensing the strategic use of states' libel laws before local juries to silence critical media coverage, the Court announced the rigorous actual malice test for libel when brought by public officials.[77]

Despite introducing this more rigorous standard, the decision to preserve a tort addressing grievances for false and defamatory statements signifies the Court's willingness to recognize and ameliorate the harm that false speech can perpetrate.[78] The *New York Times* Court protected false speech only to protect true speech.[79] This is consistent with the truth-protecting rationale for protecting false speech. For example, the Court has declared "[u]ntruthful speech, commercial or otherwise, has never been protected for its own sake."[80] Scholars characterize *New York Times* and its progeny's protection of false speech as "purely prophylactic."[81] They explain that the First Amendment "provides protection to the truth-speaker by also incidentally protecting the liar."[82] Professor Helen Norton similarly notes that the *New York Times* Court protected false statements "not because the [false] speech itself is valuable, but because government efforts to regulate such [false] speech might chill individuals' willingness to engage in valuable expression."[83]

Similarly, the Court has upheld regulations prohibiting false commercial speech because such speech deceives consumers and does not contribute to the marketplace of ideas.[84] It is also clear that laws prohibiting the making of false statements under oath, falsely shouting "fire" in a crowded theater, or engaging in fraudulent activities are justified by the need to subvert their respective real harms.[85]

In contrast, in *United States v. Alvarez*, the Court struck down the Stolen Valor Act, a statute imposing criminal sanctions on individuals falsely claiming to have received military medals, honors, or decorations, because

---

[75] *See generally id*.

[76] *Id*. at 262–64.

[77] *See generally id*.

[78] *See* Sunstein, *supra* note 71, at 891–92.

[79] *See New York Times*, 376 U.S. at 271–72 (1964) ("erroneous statement is inevitable in free debate, and it must be protected if the freedoms of expression are to have the breathing space that they need to survive"); *see also* Josh M. Parker, *The Stolen Valor Act as Constitutional: Bringing Coherence to First Amendment Analysis of False-Speech Restrictions*, 78 U. CHI. L. REV. 1503, 1511 (2011). Professor Eugene Volokh explained of *New York Times*, that the Court afforded constitutional protection to a knowingly false statement of fact because "the risk of liability for falsehoods tends to deter not just false statements but also true statements." Eugene Volokh, *Amicus Curiae Brief: Boundaries of the First Amendment's "False Statements of Fact" Exception*, 6 STAN. J. CIV. RIGHTS & CIV. LIBERTIES 343, 351 (2010).

[80] Va. State Bd. of Pharmacy v. Va. Citizens Consumer Council, Inc., 425 U.S. 748, 771 (1976).

[81] Alan K. Chen & Justin Marceau, *High Value Lies, Ugly Truths, and the First Amendment*, 68 VAND. L. REV. 1435, 1437 (2015).

[82] *Id*.

[83] Helen Norton, *Lies and the Constitution*, 2012 S. CT. REV. 161, 169 (2012).

[84] *See* Redish & Voils, *supra* note 66, at 767 ("False commercial speech, of course, serves no value in and of itself; indeed, it is reasonable to believe that it can only be harmful to society and the individuals who populate it, in a variety of ways.").

[85] S. Elizabeth Wilborn Malloy & Ronald J. Krotoszynski, Jr., *Recalibrating the Cost of Harm Advocacy: Getting Beyond Brandenburg*, 41 WM. & MARY L. REV. 1159, 1163 (2000).

the government failed to show that falsity in this context caused real harm.[86] At a public meeting, Alvarez claimed to have been awarded the Congressional Medal of Honor, which was "false" [87] and "an intended, undoubted lie."[88] The government argued that the regulation was "necessary to preserve the integrity and purpose of the Medal," therefore justifying its prohibition of the false speech compromising that integrity.[89] The Court did not find that lying about military honors was a significant harm. In a plurality opinion, Justice Kennedy, therefore, identified the act as a content-based restriction on speech and applied strict scrutiny.[90]

The Court illustrates its attitude that falsity, in and of itself, is of little value in dicta with statements like "there is no constitutional value in a false statement of fact,"[91] "[f]alse statements of fact are particularly valueless [because] they interfere with the truth-seeking function of the marketplace of ideas,"[92] and "demonstrable falsehoods are not protected by the First Amendment in the same manner as truthful statements."[93]

Yet, when the government in *United States v. Alvarez* cited these statements to argue that falsity is utterly without value and outside First Amendment protections, the Court pushed back.[94] Such pronouncements about falsity, the *Alvarez* plurality explained, were made in the context of considering "legally cognizable harm[s] associated with a false statement" such as defamation and fraud.[95] These statements, the Court explained, were not in reference to falsity alone and are therefore inapplicable to falsity devoid of a legally cognizable harm.[96]

Professors Alan Chen and Justin Marceau characterize *Alvarez* as a shift from the prophylactic justification for protecting false speech (protecting true speech) to a new approach that weighs false speech, regardless of its value, against the harm it causes.[97] Thus, a "valueless" lie "of self-promotion" about military honors that does not substantially harm is a protected falsity because the false statement did not harm.[98]

Ultimately, when regulations of false speech are permissible, it is not merely because they contain a falsity but rather because the speech's falsity presents *real harms*. All nine justices in *Alvarez* agree that "lies that cause no real harm are protected"[99]—yet, it is the presence of sufficient harms that seem to justify regulations of false speech. In sum, the Court has generally

---

[86] *See generally* United States v. Alvarez, 567 U.S. 709 (2012).
[87] *Id*. at 715.
[88] *Id*.
[89] *Id*. at 716.
[90] *See id*. at 726. Justice Breyer wrote a concurring opinion calling for the application of intermediate scrutiny stressing that the Stolen Valor Act's implication on free speech should be balanced against the government's interest in regulating the false speech. *Id*. at 729–31 (Breyer, J., concurring in judgment).
[91] Gertz v. Robert Welch, Inc., 418 U.S. 323, 340 (1974).
[92] Hustler Magazine, Inc. v. Falwell, 485 U.S. 46, 52 (1988).
[93] Brown v. Hartlage, 456 U.S. 45, 60 (1982).
[94] *See generally* United States v. Alvarez, 567 U.S. 709 (2012) (plurality).
[95] *Id*. at 719.
[96] *Id*.
[97] *See* Chen & Marceau, *supra* note 81, at 1435, 1437, 1452.
[98] *See id*. at 1452–53.
[99] *Id*. at 1480. Chen and Marceau explain that six justices require a legally cognizable harms to justify regulations of falsity and three dissenting justices "recognized that only those lies that 'inflict real harm and serve no legitimate interest' fall outside the protection of the First Amendment." *Id*.

upheld regulations of false speech where the speech's falsity presents real and concrete harms but declines to regulate speech merely because it is false.

Hateful falsity is a type of falsity that *actually* harms by causing antisocial behavior.[100] The harm of antisocial behavior exceeds libel's harm to reputation and false commercial speech's harm of consumer deception. Similar to libel and false commercial speech, the reason to regulate hateful falsity is non-censorial. Moreover, various features laid out in the sample regulation offer sufficient breathing room for true, high-value speech. By incorporating many of the same elements of the *New York Times* defamation standard, my sample regulation of hateful falsity does not impede public debate and leaves sufficient breathing space without sacrificing truth.

2. Hostility to Content-Based and Hate Speech Regulations

While the harmfulness of the falsehood encompassed in hateful falsity favors regulation, regulation must overcome additional barriers—content-based analysis and the Court's hostility to regulations of hate speech. The fundamental and almost dispositive question when considering the constitutionality of a regulation on speech is whether the regulation distinguishes between speech based on its content.[101] In 1972, the Supreme Court announced, "[a]bove all else, the First Amendment means that government has no power to restrict expression because of its message, its ideas, its subject matter, or its content."[102]

This evolved into the overarching "cardinal rule" of First Amendment free speech law[103]—laws that distinguish speech on the basis of their content, by viewpoint, or subject-matter, are defined as content-based while those that do not are defined as content-neutral.[104] The content-based dichotomy has a reputation of being the be-all-end-all of free speech law because of its results.[105] The cardinal rule has earned this reputation because of the drastically different scrutiny applied to content-based and content-neutral laws.

The Court reviews content-neutral laws under the more deferential intermediate scrutiny, upholding regulations as long as they "further[] an important or substantial governmental interest . . . unrelated to the suppression of free expression . . . [and] the incidental restriction on alleged First Amendment freedoms is no greater than is essential to the furtherance of that interest."[106] Examples of content-neutral regulations include: "[l]aws that restrict noisy speeches near a hospital, ban[s on] billboards in residential communities, limit[s on] campaign contributions, or prohibit[ions on] the mutilation of draft cards . . . ."[107]

---

[100] *See supra* Part III.

[101] Brown, *supra* note 11, at 955, 967.

[102] Police Dep't of Chi. v. Mosley, 408 U.S. 92, 95 (1972).

[103] Brown, *supra* note 11, at 954.

[104] Burson v. Freeman, 504 U.S. 191, 197 (1992) ("This Court has held that the First Amendment's hostility to content-based regulation extends not only to a restriction on a particular viewpoint, but also to a prohibition of public discussion of an entire topic."); ERWIN CHEMERINSKY, CONSTITUTIONAL LAW: PRINCIPLES AND POLICIES 978 (5th ed. 2015).

[105] *See* Turner Broad. Sys., Inc. v. F.C.C., 512 U.S. 622, 641–42 (1994) (introducing different treatment of content-based and content-neutral laws).

[106] United States v. O'Brien, 391 U.S. 367, 377 (1968).

[107] Geoffrey R. Stone, *Content-Neutral Restrictions*, 54 U. CHI. L. REV. 46, 48 (1987).

By contrast, content-based restrictions are subject to the more onerous strict scrutiny review.[108] Surviving strict scrutiny requires the government to prove that the regulation is narrowly tailored to serve a compelling government interest.[109] The most significant hurdle is not establishing a compelling government interest, but rather proving that the regulation is narrowly tailored to achieve that interest. In effect, labeling a law "content-based" is the kiss of death because of the strict scrutiny review that inevitably follows.[110] The Court frequently reminds us that content-based laws are "presumptively invalid"[111] and "it is the rare case in which we have held that a law survives strict scrutiny."[112] Only a handful of content-based regulations on speech have actually survived judicial review.[113]

Under the cardinal rule, the Court will indiscriminately apply strict scrutiny to any regulation that distinguishes speech based on content, regardless of whether the law seeks to suppress an idea or address real harms. This indiscriminate application equates censorial regulations that aim to suppress ideas with non-censorial regulations in which government exercises its police powers to provide for the general welfare. Professor Rebecca Brown argues that content-based analysis screens for the wrong concerns and hinders government's ability to exercise its police power to protect for the general welfare and ensure that the constitutional rights of all are protected.[114]

Consequently, by prioritizing the screening of the wrong things (content-based distinctions over censorship), speech that ought to be free and left unregulated can still evade the Court's protection while regulations on speech that serve no censorship purpose fail at the Court's feet.[115] For example, in *Holder v. Humanitarian Law Project*, a statute banning material support to certain foreign organizations determined to be engaged in terrorist activities prevented human rights organizations from counseling these groups on how to take advantage of lawful and peaceful remedies, petition international bodies, engage in advocacy, and construct peace

---

[108] Carey v. Brown, 447 U.S. 455, 465 (1980) ("[C]ertain state interests may be so compelling that where no adequate alternatives exist a content-based distinction—if narrowly drawn—would be a permissible way of furthering those objectives."); *see* Barry P. McDonald, *Speech and Distrust: Rethinking the Content Approach to Protecting the Freedom of Expression,* 81 NOTRE DAME L. REV. 1347, 1363 (2006). Professor Barry McDonald explains *Carey's* reference to strict scrutiny as the origin of the application of strict scrutiny to content-based restrictions. *Id*. Since *Carey*, the Court has struck down content-based regulations under strict scrutiny in over twenty cases. *Id*.

[109] *See, e.g*., *Carey*, 447 U.S. at 465.

[110] Burson v. Freeman, 504 U.S. 191, 211 (1992) (plurality).

[111] R.A.V. v. City of St. Paul, 505 U.S. 377, 382 (1992).

[112] *Burson*, 504 U.S. at 211.

[113] *See*, *e.g*., *id*. (holding that a statute prohibiting voter solicitation and the distribution of campaign material within one hundred feet of the polling place was narrowly tailored to serve a compelling state interest); Williams-Yulee v. Fla. Bar, 575 U.S. 433, 434–36 (2015) (a judicial canon restricting judicial candidates from personally soliciting contributions served the compelling interest of "preserving public confidence in the integrity of the judiciary" in a manner narrowly tailored to serve that end).

[114] *See* Brown, *supra* note 11, at 962 (arguing that when content-based laws are the most efficient way for government, under its police power, to regulate social harms for non-suppression purposes, the regulation should not face the presumption of invalidity).

[115] *See id*. at 957–60. The rationale of the "cardinal rule"— apply strict scrutiny when a law distinguishes speech on the basis of content—is overprotection. *Id*. at 957. The theory goes that overprotecting free speech further preserves our liberties. *Id*. Brown argues that the opposite is the case: "over-protection does hurt out liberty." *Id*. By hindering government's ability to protect for the general welfare, our liberties are at risk. *Id*. at 957–58.

negotiations.[116] The Court recognized the assistance as pure speech yet nevertheless held that criminalizing this pure speech did not violate the First Amendment.[117] In reviewing this "content-based restriction on free speech," the advocates' pure speech was deemed "fungible" support to the foreign groups and thereby enabled the Court to bypass the strict-scrutiny analysis that accompanies a content-based regulation.[118] As a result, peaceful speech regarding political and international affairs could constitutionally be prohibited.

Alternatively, in *United States v. Stevens*, the commercial creation and distribution of "animal crush videos," which brutally depicted the slow death of animals by a high-heeled stiletto for the audience's sexual gratification[119] could not be prohibited without violating the First Amendment.[120] These two competing results—prohibiting political speech and protecting depictions of animal cruelty—highlight the flaws in applying the cardinal rule.[121] Faced with outcomes that seem at odds with a common-sense understanding of free speech highlight the Court's adherence to a process—identify a law as content-based and apply strict scrutiny—while overlooking the censorial motivations behind the speech restriction.

To understand how the Court would treat a seemingly content-based regulation of hateful falsity, it is important to understand the Court's treatment of hate speech and the Court's designation of "unprotected" categories of speech. The Court's hostility to content-based regulation has doomed government's ability to regulate most forms of hate speech. Hate speech is not precisely defined in First Amendment doctrine. Hate speech highlights the tension between protecting free speech and protecting citizens from harm. Some absolutists urge us to have "tough skin," as free speech requires tolerating offensive speech. Other absolutists fear the slippery slope and believe regulating hate speech cannot be done without simultaneously targeting protected speech.[122]

By contrast, those who support regulation argue that hate speech perpetrates uniquely insidious and dangerous harms, including: extensive psychological damage to the victim; harm to human dignity that "undermines the constitutional value of equality";[123] promotion of intolerance; and incitement to violence. But such theories of harm have, to date, failed to persuade the Court; instead, the Court has focused on how such regulations seek to suppress speech because of the underlying idea.[124]

---

[116] *See generally* Holder v. Humanitarian Law Project, 561 U.S. 1 (2010).

[117] *Id*. at 16–18.

[118] *See id*. at 36–38; *see also* Brown, *supra* note 11, at 959, 959 n.27.

[119] The legislative history of Section 48 of the statute informs that Congress aimed to target "crush videos." United States v. Stevens, 559 U.S. 460, 465–66 (2010).

[120] The Court struck down the statute on overbreadth grounds yet made clear strict scrutiny would apply. *See id*. at 468, 472 ("[the statute] explicitly regulates expression based on content" and "we review Steven's First Amendment challenge under our existing doctrine").

[121] *See* Brown, *supra* note 11, at 957–60.

[122] *See* CHEMERINSKY, *supra* note 104, at 1062.

[123] *Id*.

[124] See Brown, *supra* note 11, at 999–1001.

Unprotected speech refers to speech identified *for its content*,[125] and is considered outside the scope of First Amendment protection.[126] Designating categories of speech as unprotected is inherently inconsistent with the Court's own fundamental rule against content-based restrictions.[127] Regulations of unprotected speech need not overcome strict, intermediate, or any other level of heightened scrutiny. Unprotected categories of speech include: obscenity, incitement, fighting words, libel, and child pornography made with real children.[128] In 2010, the Court refused to add violent speech as a new unprotected category and suggested that the list of unprotected categories was closed.[129] The Court reiterated that "content-based restrictions . . . have been permitted, as a general matter, only when confined to the few historical and tradition categories of expression long familiar to the bar."[130]

In regulating hate speech, legislatures mostly rely on these "unprotected" categories.[131] One method that has been used is to prohibit libel perpetrated on a racial or religious group.[132] The more common method is to prohibit "fighting words" that express hate. Fighting words "by their very utterance inflict injury or tend to incite an immediate breach of the peace."[133] When introducing the doctrine, the Court in *Chaplinsky v. New Hampshire* explained: "Such utterances are no essential part of any exposition of ideas, and are of such slight social value as a step to truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality."[134] Epithets and personal abuse, the *Chaplinsky* Court explained, are not a form of communication protected by the First Amendment.[135] Based on this language one might assume that the fighting words doctrine opens the door to regulations of hate speech.[136]

The Court, however, has declined to uphold a regulation of hate speech under the fighting words doctrine.[137] In *R.A.V. v. City of St. Paul*, for example,

---

[125] *See* CHEMERINSKY, *supra* note 104, at 1037 (arguing unprotected categories "are defined based on the subject matter of the speech and thus represent an exception to the usual rule that content-based rules must meet strict scrutiny").

[126] *See* Chaplinsky v. New Hampshire, 315 U.S. 568, 571–72 (1942). Proclaiming that "it is well understood that the free speech is not absolute, at all times and under all circumstances," the *Chaplinsky* Court announced that there are "certain well-defined and narrowly limited classes of speech" including: obscenity, incitement, "fighting words," and libel. *Id*. The speech is designated as unprotected because "prevention and punishment [of such speech has] . . . never been thought to raise any Constitutional problem." *Id*.

[127] *See* CHEMERINSKY, *supra* note 104, at 1037.

[128] *See Chaplinsky*, 315 U.S. at 571–72. Child pornography made with real children has been added to the *Chaplinsky* list. New York v. Ferber, 458 U.S. 747, 764–66 (1982).

[129] United States v. Stevens, 559 U.S. 460, 469 (2010).

[130] United States v. Alvarez, 567 U.S. 709, 715 (2012).

[131] *See* CHEMERINSKY, *supra* note 104, at 1062–77.

[132] *See generally* Beauharnais v. Illinois, 343 U.S. 250 (1952) (holding an Illinois law prohibiting the dissemination of "false or malicious defamation of racial and religious groups in public places" is constitutional under libel doctrines). While still considered good law, many believe that *Beauharnais* will likely be overturned if another group libel case heads to court. *See* CHEMERINSKY, *supra* note 104, at 1063 (citing Am. Booksellers Ass'n. v. Hudnut, 771 F.2d 323 (7th Cir. 1985); Collin v. Smith, 578 F.2d 1197, 1204–05 (7th Cir. 1978)).

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

[134] *Id*. at 572.

[135] *Id*.

[136] *Cf*. R.A.V. v. City of St. Paul, 505 U.S. 377, 383–84 (1992). Justice Scalia reframed the conception of unprotected categories, stating that they are not "entirely invisible to the Constitution, so that they may be made the vehicles for content discrimination unrelated to the distinctively proscribable content." *Id*.

[137] Since announcing the "fighting words" exception in *Chaplinsky*¸ the Court has not upheld a law prohibiting fighting words. CHEMERINSKY, *supra* note 104, at 1065.

the city sought to regulate *some* fighting words by prohibiting the placement of symbols on another's property that was reasonably known to anger or alarm "on the basis of race, color, creed, religion, or gender," such as swastikas or burning crosses.[138] Writing for the Court, Justice Scalia declared that government could not make content-based distinctions of speech *within* the unprotected categories when regulating.[139] Justice Scalia wrote,"[a]ssuming, *arguendo*, that all of the expression reached by the ordinance is proscribable under the 'fighting words' doctrine, we nonetheless conclude that the ordinance is facially unconstitutional in that it prohibits otherwise permitted speech solely on the basis of the subjects the speech addresses."[140]

The Court identified two exceptions to this rule. The content regulation can stand when the basis for the law's content discrimination is "the very reason the entire class of speech at issue is proscribable" because the basis for the unprotected class has "been adjudged neutral enough" thereby relieving concerns that the regulation targets in a non-neutral way. Alternatively, when a "content-defined subclass . . . [is] associated with particular secondary effects of the speech, so that the regulation is justified without reference to the content of the . . . speech," the content-based distinction can stand.[141] That the Court found the prohibition on placing burning crosses on another's property at odds with the basis for why words that "by their very utterance inflict injury or tend to incite an immediate breach of the peace"[142] are unprotected highlight the Court's hostility to hate speech regulations.

Cognizant of the consequences of applying strict scrutiny, the Court appears to engage in a selective *ad hoc* application of content-based analysis when it instinctually believes a law ought to be saved.[143] Brown explains that, in these instances, the Court makes "moves worthy of Cirque du Soleil to avoid characterizing such regulations as content-based in the first place."[144] This acrobatic avoidance demonstrates the Justices' value judgments that it is worth protecting the public from certain harms.[145]

The regulation of hateful falsity demonstrates the problems with the cardinal rule's absolutist approach and offers the Court another opportunity to practice its acrobatics. The regulation of hateful falsity is content-based and therefore would normally trigger a strict scrutiny review. The application of a content-based framework, however, ignores the government's non-censorial purpose and motivation for the regulation—addressing antisocial behaviors that threaten society. When a regulation clearly does not target an idea but instead aims to protect its citizens from dangerous antisocial behavior, government should not be prevented from protecting society from this harm. Despite the cardinal rule's status as the bedrock principle of First Amendment law, the Court has demonstrated willingness to abandon

[138] *R.A.V.*, 505 U.S. at 380.
[139] *Id*. at 381.
[140] *Id*.
[141] *Id*. at 388–40.
[142] Chaplinsky v. New Hampshire, 315 U.S. 568, 571–72 (1942).
[143] Brown, *supra* note 11, at 956.
[144] *Id*. at 958.
[145] *Id*.

content-based analysis when it believes the harm is truly worth regulating and the regulation is non-censorial.[146] This suggests that the Court may find a way to allow regulation of hateful falsity.

3. First Amendment Theories

Regulating hateful falsity adheres to the theoretical underpinnings of the First Amendment. Scholars identify four rationales for protecting free speech.[147] First, free speech enables self-governance and our democratic system; we self-govern by voting, which requires open public discourse and political dissent.[148] Second, free speech has a truth-seeking function, often summarized by Justice Holmes' famous "marketplace of ideas" metaphor.[149] By allowing speech and ideas to compete in the marketplace, truth flourishes.[150] In classic laissez-faire fashion, competition cures all; more speech is the cure for bad speech.[151] Third, free speech enables our personal autonomy by allowing us to define ourselves through our expression.[152] Finally, free speech is "integral to tolerance, which should be a basic value in our society."[153]

Hateful falsity undermines First Amendment values. Regulation is, therefore, necessary to advance the above-enumerated values. By reducing the incidence of hateful falsity, government would advance tolerance. As the data shows, exposure to hateful falsity desensitizes the audience and reduces sympathy for the victim of the hate speech. Here, unregulated hateful falsity is antithetical to tolerance.

Self-government and individual autonomy are also undermined by hateful falsity. Another consequence of hateful falsity's antisocial behavior is an increased distrust in institutions, feelings of powerlessness, and depressed civic and democratic engagement.[154] Psychologists Jolley and Douglas explain that civic engagement is decreasing worldwide and the increased exposure to conspiracy theories is partially responsible.[155] Conspiracy theories relating to government are associated with feelings of powerlessness which can lead individuals to believe their actions are inconsequential, thereby reducing their intention to vote in elections.[156] This is worthy of pause and repetition—exposure to hateful falsity sparks an unconscious psychological inclination to self-disenfranchise and refrain from democratic engagement, thereby hindering society's ability to self-govern. Given that self-government requires "meaningful deliberation"—the process of forming a public opinion requires protection.[157] Alleviating the

---

[146] *See generally id*.
[147] CHEMERINSKY, *supra* note 104, at 969–70.
[148] Alexander Meiklejohn, *The First Amendment Is an Absolute*, 1961 SUP. CT. REV. 245, 255 (1961).
[149] Abrams v. United States, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting).
[150] *Id*.
[151] *See* Whitney v. California, 274 U.S. 357, 375–77 (1927) (Brandeis, J., concurring) ("the remedy to be applied is more speech, not enforced silence").
[152] CHEMERINSKY, *supra* note 104, at 973.
[153] *Id*. at 974.
[154] *See* Jolley & Douglas, *supra* note 45, at 37; van der Linden, *supra* note 4, at 173.
[155] *See* Jolley & Douglas, *supra* note 45, at 41.
[156] *See id*.
[157] Chen & Marceau, *supra* note 81, at 1473–74 (citing Robert Post, *Participatory Democracy and Free Speech*, 97 VA. L. REV. 477, 483 (2011)).

burdens and obstructions to this process promotes the self-governance rationale underlying free speech. Moreover, hateful falsities' manipulative nature violates our individual autonomy. The consumption process of hateful falsities undermines the listener's autonomy by manipulating their psychology. Thus, the threat to self-government and individual autonomy is far greater when this speech is left unregulated. Given our democratic social contract and our fundamental commitment to the democratic process, allowing government to prevent these harms is surely justified.

Finally, Justice Brandeis' coined marketplace of ideas metaphor is at odds with hateful falsity. Justice Brandeis embraced an open marketplace where all ideas can compete clearing the path for truth. This, however, is incompatible with hateful falsity. Given the internet's echo chambers and filter bubbles and the isolated hubs of conspiracy theories, the blockade on outside information distorts the marketplace and causes market failure. Without an adequate marketplace, the truth-seeking rationale of the marketplace collapses.

This market failure is caused because the internet, social media, and technology offer isolation and a limited supply of information. Through the internet, people create their own echo chamber.[158] The infinite amount of content, sources, and outlets allow individuals to cherry-pick what information they are exposed to.[159] Professors Alstyne and Brynjolfsson explain that when individuals can "screen out material . . . [they] insulate themselves from opposing points of view and reinforce their biases. Internet users . . . thus become less likely to trust important decisions to people whose values differ from their own."[160]

"Filter bubbles" further exacerbate the problem. To increase user engagement and time spent on a site, the site learns its users' interests and ideologies over time and then employs algorithms to supply users with content adhering to those interests and ideologies.[161] Consequently, exposure to different perspectives and ideas diminishes. People can "beat the algorithms" only if they actively seek out the alternative viewpoints, however, ordinary use of the site subjects users to the filter bubbles.[162] This further tightens the homogeneity of the echo chamber; as a result, even though people are reading about the same issues, they "are not having the same conversations."[163]

Naturally, echo chambers and filter bubbles reduce individuals' access to truthful information, and hateful falsity flourishes in such environments. Individuals' exposure to certain information—the choice of what people read and listen to—impacts their susceptibility to conspiracy theories.[164]

---

[158] David R. Grimes, *Echo Chambers Are Dangerous – We Must Try to Break Free of Our Online Bubbles*, GUARDIAN (Dec. 4, 2017), https://www.theguardian.com/science/blog/2017/dec/04/echo-chambers-are-dangerous-we-must-try-to-break-free-of-our-online-bubbles.

[159] *Id*.

[160] Marshall Van Alstyne & Erik Brynjolfsson, *Global Village or Cyber-Balkans? Modeling and Measuring the Integration of Electronic Communities*, 51 MGMT. SCI. 851, 865–866 (2005).

[161] *See the Reason Your Feed Became an Echo Chamber – And What to Do About It*, NPR: ALL TECH CONSIDERED (July 24, 2016, 6:01 AM), https://www.npr.org/sections/alltechconsidered/2016/07/24/486941582/the-reason-your-feed-became-an-echo-chamber-and-what-to-do-about-it.

[162] *Id*.

[163] *Id*.

[164] Sunstein & Vermeule, *supra* note 22, at 211–12.

Conspiracy theories breed in environments where the quantity of relevant information available is low.[165] When this occurs, individuals are learning about "events" with limited information about the event.[166] Sunstein and Vermeule analogize this to extremism. Extremists grow from a lack of information and "their extremists views are supported by what little they know."[167] It is ironic that in an age when the world's knowledge is available at our fingertips people actively confine themselves to low information environments. This, alongside the internet's echo chambers and filter bubbles, prevent the marketplace from offering good speech to cure bad speech.

We fight for free speech protection because of the belief that free speech promotes tolerance, furthers self-government, protects individual autonomy, and enables a marketplace of ideas where truth can flourish. Yet, when certain speech serves to undermine, hinder, and prevent the fruition of these values, a reckoning of how we choose to handle such speech is necessary. Our continued tolerance of hateful falsity forces a choice: do we prioritize this speech for the sake of protecting *more* speech or will we choose our First Amendment values?

4. Strict Scrutiny

In spite of a non-censorial reason to regulate hateful falsity, hateful falsity's real harm of antisocial behavior, its false nature, and its undermining of First Amendment values, the Court may still invoke the cardinal rule and apply strict scrutiny. As stated above, when a regulation is subject to strict scrutiny, the government must demonstrate that it has a compelling interest in regulating the speech and its regulation is necessarily tailored to achieve those ends. In *R.A.V.*, the Court accepted protecting against the social harms inflicted by bias-motivated threats to public safety as a compelling interest.[168] Further, false statements of fact have been found to "cause tangible social harm"; for example, "unnecessarily alarming people might cause panic, leading to physical injuries."[169] Additionally, given the harms described above, the Court is likely to find the abatement of antisocial behavior as a compelling government interest.

The regulation of hateful falsity is narrowly tailored to tackle the nationwide harm that hateful falsity inflicts. In light of the marketplaces' failures and an inability for private actors to adequately respond, the burden falls on government to act. The evidence of how quickly hateful falsity spreads, its cognitive consumption process, and the isolated hubs in which it grows demonstrate that a less restrictive regulation would insufficiently address the society-wide antisocial behavior attributed to hateful falsity.

A prohibition on hateful falsity is, therefore, necessary and represents the least restrictive means of regulation. Other means would fail. For

---

[165] *See id*.
[166] *See id*.
[167] *See id*.
[168] *See* R.A.V. v. City of St. Paul, 505 U.S. 377, 395 (1992) (finding that ensuring minority groups historically subject to discrimination be able to live in peace is a compelling government interest despite striking down the law for failing to narrowly achieve that end).
[169] Chen & Marceau, *supra* note 81, at 1440.

instance, attempts to counter false speech with truthful factual speech has and will continue to be unsuccessful because of the aversion to and blockade of truth in these forums, and because denial furthers conspiracy theories.[170] Thus, a systematic practice by government to counter hateful falsity will fail and potentially make matters worse.[171] Government action will lend credence and legitimacy to the falsehood. Some might see the government's rebuttal as an indication that the theory is credible since it justified government attention. This expands the audience that finds the falsity credible. Government counter-speech can also serve as further evidence of the cover-up and appear as if powerful forces are at work as denial is internalized as part of the coverup.[172] Furthermore, *ad hoc* discretion over which speech to counter starts to look like censorship, which is susceptible to viewpoint discrimination. Given that a prohibition is necessary to address the society-wide implications of the antisocial behavior caused by hateful falsity, the regulation must take the form of a prohibition.

Pursuing options in tort law is also inadequate because tort law fails to capture the collective nature of the social harm of hateful falsity. Libel and intentional infliction of emotional distress claims focus on an individualized theory of the harm. These claims would remedy individual harms (e.g., an individual's reputation, emotional damage from fear or harassment, etc.), but their individualized nature and instance-by-instance approach would prevent the government from addressing the real harm, society-wide harm of antisocial behavior. This is true even of group libel. Suing on behalf of a group for false statements made about the group addresses the harm hateful falsity has done to that group, and not to society. It remedies specific plaintiffs' injuries. Tort law's focus on individual harms would miss the mark.

The regulation's limiting features prevent chilling of other protected speech while ensuring it remains able to appropriately address the problems of antisocial behavior. To stop the spread of antisocial behavior, the regulation must target hateful falsity at its inception by preventing its public dissemination because conspiracy theories are almost irrefutable once embraced and further enables additional exposure. Prohibiting the public dissemination of hateful falsity strikes a balance between over-inclusiveness and under-inclusiveness. It avoids regulating speech that is not responsible for the crux of the social harm while simultaneously ensuring that it targets the speech that causes antisocial behavior. The falsity, materiality, and *mens rea* requirements likewise strike those balances—targeting the speech that causes antisocial behavior, while ensuring it does not overregulate speech that does not.

---

[170] Sunstein & Vermeule, *supra* note 22, at 221–22,

[171] *Cf*. *Id*. at 219. Sunstein and Vermeule propose that the government penetrates these hubs and pumps in truthful speech. *Id*. Their model is aimed at cutting between the supply of conspiracy theory and its audience. *Id*. They propose that the government engages in "cognitive infiltration of extremists groups" by pumping truth into the hub to undermine the rationale the audience relies on to accept the conspiracy theory as true. *Id*.

[172] *See id*. at 219, 221–22.

V. CONCLUSION

Regulation of hateful falsity is not justified by the fact that it offends, humiliates, and demeans its victims, as it would seek to suppress an idea. The First Amendment protects even abhorrent ideas. Rather, regulation of hateful falsity is justified because it harms society by causing widespread antisocial behavior. As individuals become more antisocial, the harms aggregate. Paranoia, aggression, verbal assault, violence, and civic disenfranchisement are just some of the consequences. It is for these non-censorial reasons that we need to regulate hateful falsity. The severity of the harm and these non-censorial motivations permit government action.

To conclude that the First Amendment prevents the government from protecting society against these real harms as a result of indiscriminate treatment of content-based regulations would hinder government's core function: to protect our general welfare.[173] Government's ability to shield its citizens from such harms is the essence of our democratic social contract. It is why we cede power to government, so it may exercise its power to secure our rights to life, liberty, and the pursuit of happiness.[174]

[173] *See generally* Brown, *supra* note 11.
[174] *See* THE DECLARATION OF INDEPENDENCE para. 2 (U.S. 1776).