DOES HATE SPEECH VIOLATE FREEDOM OF THOUGHT?

Lucas Swaine

I. INTRODUCTION

Hate speech has become highly salient for theories of free speech and for understandings of the proper regulatory scope of democratic institutions. The continuing presence of hate speech precipitates disputes over whether it should be legally restricted and, if so, how far government may go in regulating it. The mechanisms that government might employ, to regulate hate speech, are also highly controversial, as are justifying reasons for regulating hate speech in the first place. On the latter front, various harms have been attributed to hate speech. These include claims to the effect that hate speech is libelous,[1] that it causes social disruption and unrest,[2] that it incites hatred or abuse,[3] that it constitutes or causes psychological wounding or other adverse psychological effects,[4] and that hate speech damages one's dignity, reputation, or standing as a social and political equal.[5]

I aim to advance scholarly discussion of these issues by bringing hate speech into conversation with the value of freedom of thought. In this Article, I argue that hate speech can violate freedom of thought, independent of other harms that hate speech may constitute or precipitate. In the course of argument, I provide a framework for identifying, classifying, and weighing prospective freedom-of-thought infringements. My treatment supplies new and different ways to think about regulating hate speech, and it offers fresh approaches for upholding commitments to freedoms of

---

[1] *See* Beauharnais v. Illinois, 343 U.S. 250 (1952); THOMAS EMERSON, THE SYSTEM OF FREEDOM OF EXPRESSION 391–99 (1970); NADINE STROSSEN, HATE: WHY WE SHOULD RESIST IT WITH FREE SPEECH, NOT CENSORSHIP 44–46 (2018); *see also* JEREMY WALDRON, THE HARM IN HATE SPEECH (2012); *see generally* CATHARINE MACKINNON, ONLY WORDS 47–52, 60–61 (1993).

[2] *See* Tariq Modood, *Hate Speech: The Feelings and Beliefs of the Hated*, 13 CONTEMPORARY POLITICAL THEORY 104, 105 (2014); WALDRON, *supra* note 1, at 103–04.

[3] *See* Simon Thompson, *Understanding and Regulating Hate Speech: A Symposium on Jeremy Waldron's 'The Harm in Hate Speech'*, 13 CONTEMP. POL. THEORY 88, 88, 91, 103, 105 (2014); *cf.* C. Edwin Baker, *Autonomy and Hate Speech*, *in* EXTREME SPEECH AND DEMOCRACY 139, 146–49 (Ivan Hare & James Weinstein eds., 2010).

[4] *See* MARI J. MATSUDA ET AL., WORDS THAT WOUND: CRITICAL RACE THEORY, ASSAULTIVE SPEECH, AND THE FIRST AMENDMENT (1993); MACKINNON, *supra* note 1, at 100; RICHARD DELGADO & JEAN STEFANCIC, UNDERSTANDING WORDS THAT WOUND 12–15 (2004); *see also* STROSSEN, *supra* note 1, at 123–27, 151–53, 171–73; *see also* Eric Barendt, *What Is the Harm of Hate Speech?*, 22 ETHICAL THEORY AND MORAL PRAC. 539, 539–41 (2019).

[5] *See generally* WALDRON, *supra* note 1.

speech and of thought, and for supporting other foundational democratic rights and liberties, as well.

I begin by supplying a working conceptualization of hate speech, following which I consider a variety of putative harms associated with speech of that kind. I then offer a brief account of freedom of thought, noting its importance as a basic liberty, outlining its relationship with freedom of speech, and pointing out different ways in which freedom of thought can be infringed. I provide a framework for identifying speech-based freedom-of-thought violations, and I argue that hate speech can violate freedom of thought based on the degree to which the speech is disturbing, importunate, or insidious. The framework that I furnish is supported by examples and it tracks reasonable concerns about emotional injury and psychic wounding articulated by Critical Race Theorists. I consider prospects for regulating hate speech based on the arguments that I provide, prior to concluding.

In this Article, I focus largely on cases and examples from the United States; the argument that I generate, however, is broadly applicable and it is transportable to other liberal-democratic contexts. To be clear, I shall not aim to provide an overarching treatment of all of the harms or concerns associated with hate speech. Nor do I propose to engage in detailed analysis of the degree to which freedom-of-thought protections may resonate with different countries' existing constitutional principles and commitments. I provide instead a philosophical treatment of rights and freedoms that people should be acknowledged to have, even if the respective values are not clearly expressed in some countries' constitutional language. I make my case by building upon existing scholarship, in these respects, and appreciating that one finds differing understandings of hate speech in different democratic political and legal systems, and across theories of hate speech.

II. WHAT IS HATE SPEECH?

Consider first the nature of hate. Whereas hatred is in the running for being the "most destructive affective phenomenon" in human history,[6] some scholars have expressed puzzlement over its being a comparatively "underresearched topic."[7] Social-science research testifies to significant differences of opinion on hate's nature. Hatred has been described variously as an emotion, as an emotional attitude, as a generalized form of anger, as a kind of judgment, and as a motive, among other depictions.[8]

---

[6] Agneta Fischer et al., *Why We Hate*, 10 EMOTION REV. 309, 309 (2018); Eran Halperin et al., *In Love with Hatred: Rethinking the Role Hatred Plays in Shaping Political Behavior*, 42 J. APPLIED SOC. PSYCH. 2231, 2232 (2012).

[7] Fischer et al., *supra* note 6, at 309–10.

[8] *See id.*

I proceed with a basic understanding of "hate" as "intense or passionate dislike." Hatred is in this sense a negative emotional phenomenon[9] that involves feelings of hostility or antipathy toward something. This definition fits with common language, accommodating the wide variety of possible objects of hate that one sees referenced in everyday discussion. People hate other individuals; but one might also hate groups, activities, places, foods, institutions, cultures, races, or oneself. The elemental view of hatred as intense or passionate dislike fits with these familiar prospects. The definition has the benefit of being sufficiently broad to allow that hatred might be construed as a more immediate or long-term emotion,[10] and it accommodates the sense that hatred may function as an element of a "self-defense system" predicated on a desire to get rid of a hated object.[11]

I take "speech," for purposes of argument, simply to mean "expressive action."[12] This understanding is suitably broad to allow a wide variety of forms of expression to be included under the rubric of speech, including spoken utterances, writing, symbols, and symbolic gestures.[13] It is important not to limit speech, or the freedom of it, just to spoken words. Some forms of conduct, such as flag-burning, have been acknowledged as highly expressive actions, even though no words need be uttered during the conduct itself.[14] Cross-burning is another case in point: the United States Supreme Court treated cross-burning as speech in *R.A.V. v. City of St. Paul* and in *Virginia v. Black*, respectively.[15] As Cass Sunstein reflects, it would be difficult to maintain that flag-burning is a form of speech but that cross-burning is not.[16] To this one can add that the ability to generate expressive acts need not be restricted only to individuals; groups and organizations could at times qualify as expressive actors, too. Owen Fiss submits that a commitment to the protection of institutional autonomy

---

[9] *Id.* at 310; *see, e.g.*, Robert J. Sternberg, *A Duplex Theory of Hate: Development and Application to Terrorism, Massacres, and Genocide*, 7 REV. GEN. PSYCH. 299, 304–24 (2003).

[10] Fischer et al., *supra* note 6, at 317.

[11] *Id.*; Robert Post, *Hate Speech, in* EXTREME SPEECH AND DEMOCRACY 123, 123–36 (Ivan Hare & James Weinstein eds., 2010).

[12] *See* Thomas Scanlon, *A Theory of Freedom of Expression*, 1 PHIL. & PUB. AFFS. 204, 206 (1972).

[13] *See generally* Timothy Zick, *Cross Burning, Cockfighting, and Symbolic Meaning: Toward a First Amendment Ethnography*, 45 WM. & MARY L. REV. 2261 (2004); *see also* Ryan W. Davis, *Symbolic Values*, 5 J. AM. PHIL. ASS'N 449 (2019).

[14] *E.g.*, Texas v. Johnson, 491 U.S. 397 (1989); United States v. Eichman, 496 U.S. 310 (1990).

[15] R.A.V v. City of St. Paul, 505 U.S. 377 (1992); Virginia v. Black, 538 U.S. 343 (2003).

[16] CASS SUNSTEIN, THE PARTIAL CONSTITUTION 245–46 (1998); *see generally* Chris L. Brannon, Note, *Constitutional Law – Hate Speech – First Amendment Permits Ban on Cross Burning When Done with the Intent to Intimidate*, 73 MISS. L.J. 323 (2003); *see also* Zick, *supra* note 13 (2004).

would have to acknowledge that "organizations have viewpoints," and that those viewpoints are not necessarily reducible to the views of any one particular individual.[17]

Thomas Scanlon has argued that in order to categorize any act as expressive, it is sufficient to show that the act in question is "linked with some proposition or attitude" that it is intended to convey.[18] To this it is sensible to add that expressive acts[19] need not transmit any definite propositional content in order to qualify as speech. More artistic modes of expression, for example, may be ambiguous or unclear in that regard. It can be difficult to know exactly what attitudes or propositional content may or may not be connected with artistic works.[20] And yet those works can be highly expressive, even when it is difficult to understand just what they say, or when what they mean to convey is obscure.

Hate speech is a narrower construct devised with the purpose of being at least potentially suitable for inclusion in constitutional democracies' legal codes. While many conceptions of hate speech exist,[21] a basic working definition will suffice. I shall define "hate speech" as "an expressive act that communicates intense or passionate dislike of individuals or groups, based on ascriptive identity factors of those persons." This ecumenical understanding is compatible with the United Nations' recent characterization of hate speech,[22] and it is suitable for the purposes of considering the relationship between hate speech and freedom of thought.

---

[17] Owen Fiss, *Free Speech and Social Structure*, 71 IOWA L. REV. 1405, 1411 (1986); *cf.* CHRISTIAN LIST & PHILIP PETTIT, GROUP AGENCY: THE POSSIBILITY, DESIGN, AND STATUS OF CORPORATE AGENTS (2011); *see also* LUCAS SWAINE, ETHICAL AUTONOMY: THE RISE OF SELF-RULE 37 n.20 (2020) (discussing the possibility of collective or compound entities qualifying as agents, and suggesting that the qualities of such agents' actions may not be "fully reducible to those of individual persons").

[18] Scanlon, *supra* note 12, at 206.

[19] *Id.* (I take expressive acts to include not only verbal acts, but also "displays of symbols," "demonstrations," "assassinations," and so forth, as Scanlon describes); *see also* JOHN HART ELY, DEMOCRACY AND DISTRUST: A THEORY OF JUDICIAL REVIEW 113 (1980); *see generally* Zick, *supra* note 13 (2004).

[20] *See, e.g.*, Amy Adler, *What's Left?: Hate Speech, Pornography, and the Problem for Artistic Expression*, 84 CALIF. L. REV. 1499, 1541–44 (1996) (noting that some works by American photographers Andres Serrano and Robert Mapplethorpe appear to be ambiguous in the messages they communicate).

[21] *See* STROSSEN, *supra* note 1, at xxiii, 1, 11–12 (2018).

[22] *See* ANTÓNIO GUTERRES, UNITED NATIONS SECRETARY-GENERAL, UNITED NATIONS STRATEGY AND PLAN OF ACTION ON HATE SPEECH 2 (2019), https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20Plan%20of%20Action%20on%20Hate%20Speech%2018%20June%20SYNOPSIS.pdf (describing "hate speech" as "any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor").

The idea that hate speech centers on "ascriptive identity factors" fits with the United Nations' notion that hate speech attacks people based on such qualities as their religion, ethnicity, nationality, race, or gender.[23] The United Nations maintains that hate speech undermines people "on the basis of who they are."[24] This suggests that hate speech is different from merely pejorative or censorious speech regarding people's actions. The working definition thereby allows that it would not be hate speech to castigate people for breaking just and legitimate laws, or for expressing intense dislike of rights-violations. But it would count as hate speech to communicate intense dislike of a minority religious community, based on identity factors ascribed to the group or its religious practices. The basic identity traits on which hate speech is predicated may be presumed to be orthogonal to issues on which people may fairly be criticized.

This conception of hate speech squares with the observation that hate speech often takes form in group-level claims that apply to individual people, as one sees reflected in an assortment of legal understandings. For instance, the *Canada Criminal Code* describes the relevant targets of hate speech as "identifiable group[s],"[25] and the working definition is compatible with that formulation. It may also be noted that the conceptualization neither includes definitionally, nor requires otherwise, any incitement, outward threats, or encouragement of violence against others. Legal definitions of "hate speech" sometimes include mention of such elements, but one can usefully separate incitement, threats, and similar such expressions or behaviors from a more fundamental understanding. Similarly, while hate speech may at times include slanderous or libelous claims, it need not always do so, at least not on the working understanding that I furnish here.

This conception also does not predicate hate speech on the emotions or feelings of the speaker who expresses it, at the time of expression or otherwise. It allows that an expressive act can count as hate speech even if the party that expresses it neither intends to transmit a message of hate nor is motivated by hatred themselves. This is because expressive acts can convey hatred whether the respective speakers have any intense or passionate dislike of the respective individuals or groups to which the speech refers,[26] or even if the speaker is unaware of the meaning of the expressive

---

[23] *See id.*

[24] *Id.*

[25] *See* Canada Criminal Code, R.S.C., 1985, c. C-46 ss. 318-320(1). Section 318(4) of the Canada Criminal Code defines an "identifiable group" as "any section of the public distinguished by colour, race, religion, national or ethnic origin, age, sex, sexual orientation, gender identity or expression, or mental or physical disability." Section 319(2) makes it an offense to willfully promote "hatred against an identifiable group," so long as any such statements are not made only in private conversation and provided a handful of defenses are not established.

[26] *See, e.g.*, *Swastika T-shirt Backlash Forces Company to U-turn on Campaign*, BBC (Aug. 7, 2017) https://www.bbc.com/news/world-us-canada-40848372; *see also A Weird New Clothing Brand Want to Reclaim the Swastika:*

acts to which they give voice. Someone might unwittingly carry a derogatory epithet on a placard, for example, or post the epithet on a billboard without knowing what it means. Those forms of expression may presumably still count as hate speech, even though they may be rare, and assuming that many instances of hate speech are motivated by intense or passionate dislike of the individuals or groups in question.

III. WHAT'S WRONG WITH HATE SPEECH?

In addition to putative moral wrongs attached to hate speech, political and legal theorists have identified various harms associated with it. The concerns with hate speech are in this sense different from reasons for regulating speech in general. More general reasons for regulating speech include creating and maintaining orderliness, facilitating coordination, or simply in order to have freedom of speech itself, as Alexander Meiklejohn proposed.[27]

It is worth considering briefly some of the more prominent candidate harms attributed to, or associated with, hate speech. I proceed by conceiving of the idea of harm in a broad sense in order to avoid arbitrarily excluding from the outset a variety of theorists' concerns. It should be kept in view that the kinds of harms at issue are supposed to be those that could merit rightful involvement by political or legal institutions. In addition, the presumption is that government will be involved in normatively acceptable ways and to permissible degrees. One appreciates that there may be reasonable disagreement on the question of when governmental enmeshment is appropriate, or on kinds and degrees of state regulation, with regard to hate speech and otherwise. The analysis at hand allows for a considerable range of possible views in these respects.

Existing theories of hate speech offer a variety of reasons to hold that hate speech is harmful. These include, inter alia, claims that hate speech:

- is demeaning or degrading[28]
- is defamatory or libelous, possibly causing reputational harms[29]

We're Not Buying it, DAZED & CONFUSED (Aug. 5, 2017) https://www.dazeddigital.com/fashion/article/36981/1/a-weird-new-clothing-brand-want-to-reclaim-the-swastika.

[27] *See generally* ALEXANDER MEIKLEJOHN, FREE SPEECH AND ITS RELATION TO SELF-GOVERNMENT 18 (New York: Harper & Brothers eds., 1948), arguing that the legislature of a well-ordered society has "both the right and the duty to prohibit certain forms of speech." *Id.* at 18–19. Meiklejohn distinguishes between the abridging of speech, which he believes the First Amendment does not forbid, and abridging freedom of speech, which he thinks the First Amendment does not allow.

[28] *See* MATSUDA ET AL., *supra* note 4, at 1–15, 17–51.

[29] *See generally* MACKINNON, *supra* note 1; WALDRON, *supra* note 1, at 47–52, 60–61; *cf.* EMERSON, *supra* note 1, at 391–99; *See generally Beauharnais*, 343 U.S.; *cf.* STROSSEN, *supra* note 1, at 44–46.

- produces psychological harms or other adverse psychological effects[30]
- causes poor self-image and undermines hated groups' competencies[31]
- perpetuates subordination and other forms of wrongful treatment[32]
- degrades the value of hated groups' expressive freedoms[33]
- causes social disruption and unrest[34]
- incites hatred or abuse against targeted groups[35]
- can injure or wound people[36]

This list may not be fully comprehensive but it serves as a basic register of different harms attributed to hate speech, dividing those harms into distinct concerns. The harms on this list are identified variously across theories of hate speech, and they are marshaled as grounds for hate-speech regulation. The list includes harms that are presumably supposed to be caused by hate speech, as well as those that are constituted by it, following Eric Barendt's important distinction on this matter.[37]

It is not my purpose to assess all of these prospective harms in detail. For present purposes, the third and fifth points are notable. The third point includes the idea that hate speech causes or constitutes psychic wounding; the fifth picks up concerns regarding hostile and corrosive effects of hate speech on the public environment. These factors are helpful for illuminating ways in which freedom of thought is implicated in concerns about hate speech.

The idea that hate speech causes psychic wounding gained prominence in the 1990s, forwarded in the work of Catharine MacKinnon as well as by Critical Race Theorists. MacKinnon claims that hate speech promotes group defamation and harms its victims by causing "immediate

---

[30] *See* MATSUDA ET AL., *supra* note 4; MACKINNON, *supra* note 1; DELGADO & STEFANCIC, *supra* note 4, at 12–15; *cf.* STROSSEN *supra* note 1, at 123–27, 151–53, 171–73.

[31] *See* Richard Delgado, *Words That Wound: A Tort Action for Racial Insults, Epithets, and Name Calling*, *in* MATSUDA ET AL., *supra* note 4, at 89–110.

[32] *See* WALDRON, *supra* note 1.

[33] Frank I. Michelman, *Liberties, Fair Values, and Constitutional Method*, U. CHI. L. REV. 91, 91–114 (1992).

[34] *See* Modood, *supra* note 2, at 105; *see also* WALDRON, *supra* note 1, at 103–04.

[35] *See* Thompson, *supra* note 3, at 88; *cf.* Criminal Justice and Immigration Act 2008, c. 4, Part 5 (UK); *cf.* Baker, *supra* note 3, at 139–57, 146–49.

[36] *See* Delgado, *supra* note 31; Charles R. Lawrence III, *If He Hollers Let Him Go: Regulating Racist Speech on Campus*, *in* MATSUDA ET AL., *supra* note 4, 53–88; DELGADO & STEFANCIC, *supra* note 4, at 12–13.

[37] *See* Barendt, *supra* note 4, at 540–41.

psychic wounding."[38] In their book *Words That Wound*, Critical Race Theorists supply parallel arguments, contending that racial epithets and similar language cause deep psychological damage.[39] Charles Lawrence suggests that racial epithets cause "deep emotional scarring," for example, adding that "psychic injury" is no less injurious than a physical assault.[40] Richard Delgado adds a related point about the psychological effects of racial hate speech. He argues that because minority children continually hear racist messages, they "come to question their competence, intelligence, and worth."[41] Critical Race Theorists deploy these arguments to call for extensive and comprehensive regulation of hate speech; they contend that no one should have the right to humiliate or degrade anyone else.[42] They seek a constitutional community in which "freedom" no more "implicate[s] a right to degrade and humiliate another human" than it implies "a right to do physical violence to another or a right to enslave another or a right to economically exploit another in a sweatshop, in a coal mine, or in the fields."[43]

Critical Race Theorists maintain that speakers deploy hate speech to "ambush," "terrorize," "humiliate" or "degrade" others.[44] It is not obvious, however, that it ought to be necessary that someone actually feel degraded, humiliated, ambushed, or terrorized in order for an expressive act to count as hate speech, or to satisfy criteria for governmental regulation of the speech in question. The concern here is that people may experience hate speech, even hate speech targeted directly at them, but not feel pained or bothered by it. There are various ways in which this can occur. Someone might have become accustomed to hearing particular kinds of hate speech expressed toward them. A related possibility is that an individual may not experience hate speech as speech directed towards her because she disassociates herself psychologically from the hated traits identified in the expressive acts.[45] Another possibility is that those subjected to hate speech may not experience the speech as harmful because they have come to believe that the qualities in question should be hated.[46] Each of these points applies to people experiencing hate speech regarding race, nationality, religion, identity, gender, sexuality, and so on. These considerations provide

---

[38] MACKINNON, *supra* note 1, at 100.

[39] *See* MATSUDA ET AL., *supra* note 4.

[40] Lawrence, *supra* note 36, at 74; *see id.* at 72–76.

[41] Delgado, *supra* note 31; DELGADO & STEFANCIC, *supra* note 4, at 14–15.

[42] MATSUDA ET AL., *supra* note 4, at 1, 15.

[43] *Id.* at 15.

[44] *Id.* at 1.

[45] *Cf.* FRANTZ FANON, BLACK SKIN, WHITE MASKS *passim* (Richard Philcox trans., Grove Press rev. ed. 2008); DELGADO & STEFANCIC, *supra* note 4, at 14–15.

[46] *Cf.* TONI MORRISON, THE BLUEST EYE (2007); Shara Sand, *Coming Out, Being Out: Reconciling Loss and Hatred in Becoming Whole*, 20 PSYCHOANALYSIS, CULTURE & SOC'Y 250 (2015).

at least presumptive reason to hold that someone need not necessarily be pained by an expressive act in order for it to count as hate speech, or to have the expression rise to a level meriting regulation.[47]

Jeremy Waldron builds upon these ideas, extending his concerns to speech acts that "express profound disrespect, hatred, and vilification for the members of minority groups."[48] He emphasizes the toxic atmosphere that hate speech can create. Waldron's central concern is that hate speech reduces peoples' assurance of dignified treatment as equal members of society.[49] At a public level, he argues, all citizens should be able to be confident that they will be treated with dignity, with everyone understood to be a social equal in good standing.[50] This assurance is supposed to be part of the dignity that each person merits under the law.[51] Hate speech creates a hostile environment that undermines that assurance: it attacks people's reputations, tarnishing people's dignity and charging that they are unworthy of equal citizenship.[52] Spoken and other more ephemeral forms of hate speech might not properly be banned, Waldron concludes, but written hate speech may legitimately be proscribed when it undermines the shared elements of "status, dignity, and reputation" in the socio-political environment shared by all.[53]

Waldron's treatment is prescient in important respects. His arguments presage current movements to rid the public environment of monuments and statues, flags, and other "enduring," "semipermanent," or "permanent" forms of what many take to be hate speech.[54] Waldron's theory does leave some key questions unanswered, however. First, apart from concerns over just what the conceptual and normative elements of "assurance" happen to be, there remains some puzzlement over exactly what it would mean for hate speech to injure one's dignity.[55]

Second, it is unclear why merely temporary forms of hate speech should not be regulated, whether one accepts Waldron's case regarding the proscription of more enduring expressions. Those speech acts might be less permanent than the written word, but even short-lived expressions of hatred may plausibly constitute or cause harm,[56] and, as I shall argue, the harms at issue could warrant regulation of the respective expressive acts.

---

[47] *Cf.* C. Edwin Baker, *Harm, Liberty, and Free Speech* 70 S. CALIF. L. REV. 979, 986–93 (1997).

[48] WALDRON, *supra* note 1, at 27. *See also id.* at 37.

[49] *Id.* at 4, 5, 16, 69, 81–104, 107, 108, 116, 165–72, 220, 232–33.

[50] *Id.* at 5, 47, 142, 233.

[51] *Id.* at 86, 92, 218–20. *Cf. id.* at 166.

[52] *Id.* at 39–41, 47, 58–59, 61, 149, 165, 171.

[53] *Id.* at 47. *See also id.* at 27, 37, 39, 45, 60, 61, 85, 94–96, 139, 165. *Cf. id.* at 117; DELGADO & STEFANCIC, *supra* note 4, at 32, 143.

[54] WALDRON, *supra* note 1, at 33, 37–39, 45, 59, 69, 72–74, 99. *Cf. id.* at 116.

[55] *See* Barendt, *supra* note 4, at 542. *Cf.* ERIC HEINZE, HATE SPEECH AND DEMOCRATIC CITIZENSHIP (2016).

[56] *See generally* Barendt, *supra* note 4.

For example, even though it may be fleeting, spoken hate speech can be expressed to people repeatedly, communicating harmful messages over and over again. Speech can also be difficult to avoid—and harder to bypass or to ignore than written words—when it takes form as spoken utterances or as loud expressive conduct. The permanence of an expression of hatred should not be considered criterial for determining whether government may rightfully regulate it as hate speech.

Third, episodes of hate speech can occur in relative isolation, removed from the broader social sphere. It is not clear why government should not, or may not, regulate these more isolated occurrences, whether the hate speech is "enduring" and if it is detached from the social climate that Waldron describes.[57] Numerous instances of hate speech appear to be separated from wider social discourse in this way; this seems true especially for spoken or visual episodes of hate speech that do not persist in the manner with which Waldron is concerned.

Fourth, it is plausible that some particular instances of hate speech, written or unwritten, isolated or within broad social spaces, may in many cases not contribute to the overall social environment in which they are expressed. Not every expressive act reasonably can be expected to count as an addition to the overall atmosphere, just as not every weather event contributes to, or is indicative of, a region's climate. Hate speech can be localized and very personal. It may be oriented at individuals or small groups, and it can be directed toward local organizations and neighborhood institutions. It need not affect the social environment in ways that widely circulating publications, large public displays, televised or streamed protest marches, and similar expressive acts are inclined to do. What is more, supposing for sake of argument that one knew that a particular instance of hate speech would not degrade the broader social environment, it neither follows logically, nor does it seem reasonable to stipulate, that the expressive act could not merit regulation or proscription by government. Indeed, victims of disconnected episodes of hate speech could have the assurance of their public-level dignity and equality reasonably guaranteed, but nevertheless be burdened by hate speech coming from various corners, in less visible ways. A more complete account of hate-speech regulation should cover these environmentally noncontributory forms of hate speech.

IV. FREEDOM OF SPEECH AND FREEDOM OF THOUGHT

To this point, I have laid out a basic conception of hate speech and considered notable theoretical contributions with respect to what the harms in hate speech happen to be. I turn now to offer a brief account of freedom of thought, to facilitate consideration of how, or whether, hate speech can violate it.

---

[57] *See* WALDRON, *supra* note 1, at 37, 38, 45, 72. *Cf. id.* at 116.

Freedom of thought has been praised in liberal political theory, acknowledged in notable statements by such authors as Wilhelm von Humboldt, John Stuart Mill, and John Rawls.[58] It has also been recognized in human rights discourse, included notably in Article 18 of the 1948 *United Nations Declaration of Human Rights*[59] and in various other prominent international statements and resolutions.[60] Such laudatory comments are suggestive of the profound importance of freedom of thought. It is a freedom that merits placement among the basic liberties: freedom of thought is crucial for forming and revising one's conception of the good, it is essential for establishing a foundation for choice-making and for individual responsibility, and other central rights and liberties depend on freedom of thought in order to be workable and robust.[61] Freedom of thought proves crucial for realizing and sustaining core liberal-democratic values: people must be able to think without undue incursions or violations, in order for freedom of thought to exist.

For the purposes of analyzing freedom of thought, and to do so in a way suitable for addressing concerns regarding hate speech and related free-speech concerns, it serves to construe thinking in an unadorned way and at a basic level. I shall define "thinking" simply as "mental activity"; this allows one to include a broad range of mental phenomena under the umbrella of thinking.[62] On this understanding, thinking includes believing, reasoning, cogitation, remembering, sensing, desiring, and imagining. One may also include feeling under the rubric of thought, proceeding on

---

[58] *See* WILHELM VON HUMBOLDT, THE LIMITS OF STATE ACTION 66–69 (J. W. Burrow ed., 1993); JOHN STUART MILL, ON LIBERTY 11–13, 20, 31–33, 41, 56–71, 87, 100–01 (Elizabeth Rapaport ed., 1978); JOHN RAWLS, POLITICAL LIBERALISM 308, 334–37 (expanded ed. 2005); JOHN RAWLS, A THEORY OF JUSTICE 186–87 (rev. ed. 1999); JOHN RAWLS, THE LAW OF PEOPLES: WITH, THE IDEA OF PUBLIC REASON REVISITED 62–70, 74–75 (1999).

[59] G.A. Res. 217A (III), art. 18 (Dec. 10, 1948). *Cf.* Convention for the Protection of Human Rights and Fundamental Freedoms, art. 9, Nov. 4, 1950, Eur. T.S. No. 5, 213 U.N.T.S. 221.

[60] *See, e.g.*, G.A. Res. 2200A (XXI), International Covenant on Civil and Political Rights, at 55 (Dec. 19, 1966); International Covenant on Civil and Political Rights, art. 18, Dec. 19, 1966, 6 I.L.M. 368, 999 U.N.T.S. 171; G.A. Res. 36/55, Declaration on the Elimination of All Forms of Intolerance and of Discrimination Based on Religion or Belief, at 171, (Nov. 25, 1981). *Cf.* World Conference on Human Rights, *Cairo Declaration on Human Rights in Islam*, art. 10, 11, 18. 20, U.N. Doc. A/CONF.157/PC/62/Add.18 (Aug. 5, 1990).

[61] *See generally* Lucas Swaine, *Freedom of Thought as a Basic Liberty*, 46 POL. THEORY 405 (2018) [hereinafter Swaine (2018a)]; Lucas Swaine, *Legal Exemptions for Religious Feelings*, *in* RELIGIOUS EXEMPTIONS 74, 74–96, (Kevin Vallier & Michael Weber eds., 2018) [hereinafter Swaine (2018b)].

[62] *See* Swaine (2018a), *supra* note 61, at 411.

the understanding that "feeling" means "emotion or sentiment,"[63] and accepting that feeling is a form of mental activity.

It appears that freedom of thought can be violated in three central ways: through excessive investigation of people's thinking, by sanctioning people for their thoughts—with punishments or fines, for example—independent of how they have acted, and by subjecting people to nonconsensual and undue forms of thought modification.[64] I expand on these points elsewhere, but I take these simple notions as a reasonable basis on which to proceed: that freedom of thought is a basic liberty, that people have a right to this freedom, and that one's right to freedom of thought includes protections against excessive investigation of one's thinking, punishment for what one merely thinks, and extreme and unwarranted forms of thought-modification, respectively.

Freedom of thought and freedom of speech are mutually supportive liberties. Free speech allows people to express their ideas, to encounter new information, to discuss controversial topics with others, to participate politically, and to develop understandings across the full range of human experience.[65] Freedom of thought permits people to ponder and assess information they receive, to think through and reckon with their ideas and feelings, and to develop and employ their own determinations as active, responsible agents. Freedom of thought also gives people the prerogative to foster the very language, statements, and forms of expression that freedom of speech is dedicated to protecting. But the two freedoms display a complex relationship that features tensions as well as reciprocal support, and in some key respects freedoms of speech and thought can conflict. The conflicts are largely asymmetrical, given the respective natures of the two freedoms, with expressive acts appearing at times to grate against, even to infringe, freedom of thought.

When one considers ways in which speech might possibly run afoul of freedom of thought, it may be observed that prospective violations are unlikely to occur in the realms of excessive punishment or investigation. While the latter kinds of infringement are normally dependent on speech acts, they are not part of narrower concerns with freedom of speech and its relation to freedom of thought, per se. To put the point another way,

---

[63] *Cf.* MARTHA C. NUSSBAUM, THE THERAPY OF DESIRE: THEORY AND PRACTICE IN HELLENISTIC ETHICS 86–90, 242–44, 369–70 (1994) (discussing ancient Greek views of relationships between feelings and emotions); *cf. also* MARTHA C. NUSSBAUM, UPHEAVALS OF THOUGHT: THE INTELLIGENCE OF EMOTIONS 60 (2001) (distinguishing feelings that have "rich intentionality or cognitive content" from feelings that do not have such qualities).

[64] *See* Swaine (2018a), *supra* note 61, at 417, 420; Swaine (2018b), *supra* note 61.

[65] *Cf.* Seana Valentine Shiffrin, *A Thinker-Based Approach to Freedom of Speech*, 27 CONST. COMMENT. 283 (2011). Shiffrin argues for "a thinker-based free speech theory that takes to be central the individual agent's interest in the protection of the free development and operation of her mind." *Id.* at 287.

expressive acts, and the freedom to engage in them, are related to human abilities to investigate others' thoughts or to punish people for thoughts they are believed to have.[66] But freedom-of-thought violations with respect to wrongful investigation of thought, or in the category of punishment for thoughts, are independent concerns that may be set aside from the present analysis.

It appears that tensions between freedoms of thought and speech lie centrally in the category of thought-modification. The main concerns in this regard revolve around such issues as: (a) the kinds of changes agents may make to other's thoughts or ways of thinking, by employing expressive acts, (b) the degree to which people may permissibly modify others' thoughts, through verbal acts or other forms of expression, before they violate someone's right to freedom of thought, (c) the manner in which agents effectuate changes to people's thinking, and (d) the motivations and intentions of the parties involved. As to the nature of the changes in question, they include modifications to what someone thinks (i.e., changes to a person's beliefs, feelings, ideas, commitments, and so forth); alterations to how a person thinks, affecting the ways an individual engages in mental activity; and changes to someone's capabilities of thinking, involving modifications to mental skills and abilities that a person may possess.

By way of examples, medical procedures and psychological experiments in controlled environments prove capable of constituting freedom-of-thought violations in the category of thought-modification. Such practices have at times violated subjects' freedom of thought along with other rights and liberties that the people possessed.[67] Problematic forms of thought-modification are not limited to cases of experimentation on human beings, however. Expressive acts can also generate changes sufficient to violate freedom of thought, it appears, and one can adduce three kinds of examples to illustrate the point. First, some speech acts are *profoundly psychologically disturbing*, in terms of their visual or aural stimuli. Expressions of this variety can cause intense or extreme discomfort. They may for instance depict or describe rape, murder, torture, dismemberment, and other such acts. In so doing, it is plausible that they could violate the

---

[66] *See generally* Lucas Swaine, *Freedom of Thought in Political History*, *in* THE LAW AND ETHICS OF FREEDOM OF THOUGHT, VOLUME 1: NEUROSCIENCE, AUTONOMY, AND INDIVIDUAL RIGHTS 1, 1–25 (Marc Jonathan Blitz & Jan Christoph Bublitz eds., 2021).

[67] *See* JOHN MARKS, THE SEARCH FOR THE "MANCHURIAN CANDIDATE": THE CIA AND MIND CONTROL 73–104, 133–41, 223–24 (1979). *See also* SCOTT C. MONJE, THE CENTRAL INTELLIGENCE AGENCY: A DOCUMENTARY HISTORY 73–75 (2008); *cf.* H. KEITH MELTON & ROBERT WALLACE, THE OFFICIAL C.I.A. MANUAL OF TRICKERY AND DECEPTION 6–7 (2009). *Cf. also* Jan Christoph Bublitz, *Drugs, Enhancements, and Rights: Ten Points for Lawmakers to Consider*, *in* COGNITIVE ENHANCEMENT: ETHICAL AND POLICY CONSIDERATIONS IN INTERNATIONAL PERSPECTIVES 309, 309–28 (Fabrice Jotterand & Veljko Dubljević eds., 2016).

freedom of thought of those subjected to the speech in question. Second, *importunate expression* might plausibly infringe an individual's freedom of thought. I take importunate expression to consist of speech acts that one has not agreed to experience and which are very difficult to avoid, such that they pester an individual, in some cases confronting someone continually, even continuously, over a period of time. Examples here include the speaker who corners and harangues a person in a public area. Even if the speaker does not threaten or physically assault the person she provokes, she might violate the other party's freedom of thought if she refuses to leave the listener alone, and if she sustains her aggravating expression over a significant duration.

Third, *insidious expression* is a form of speech that has the potential to violate people's freedom of thought. This kind of speech deploys covert techniques, such as subliminal messages, to effectuate surreptitious changes in people's thinking. Propaganda and advertising are mundane examples of speech that can be insidious, and they might in some cases rise to the level of freedom-of-thought violations. Of the many forms of expression that propaganda and advertising can take, some may affect people profoundly and adversely, and in ways that they do not desire. To this one can add that individual speech acts may be more or less insidious, disturbing, or importunate: each of those qualities admits of degree. And these three qualities can be combined in single speech acts, or across multiple such acts, as well. The array of possible combinations makes it plausible that there are at least sufficient conditions for freedom-of-thought infractions, whether one can pinpoint necessary and sufficient conditions for such violations.

To be clear, the concern in the kinds of cases I describe is not just that speakers can adversely affect individuals who experience particular forms of expression. That bar is far too low to take seriously as a standard for rights-violations of freedom of thought. Some people are deeply shaken or offended by what amounts to very little. They are seriously troubled by things that should be of no concern, such as the idea that women have a right to participate in the workforce.[68] Others may be disturbed by otherwise normal visual displays[69] or they may have disproportionate aversive reactions to ordinary sounds.[70] Absent considerable qualification, expressive acts producing these kinds of effects could not reasonably be taken to

---

[68] *See* JOSHUA COHEN, PHILOSOPHY, POLITICS, DEMOCRACY: SELECTED ESSAYS 124, 131 (2009) (describing the "Fact of Easy Offense").

[69] *See* Kathleen B. Digre & K.C. Brennan, *Shedding Light on Photophobia*, 32 J. NEURO-OPTHALMOLOGY 68, 68–69, 71 ff., 76 (2013) (discussing psychiatric conditions associated with photophobia).

[70] *See* Miren Edelstein et al., *Misophonia: Physiological Investigations and Case Descriptions*, FRONTIERS IN HUM. NEUROSCIENCE (June 25, 2013), https://doi.org/10.3389/fnhum.2013.00296; *see also* Romke Rouw & Mercede Erfanian, *A Large-Scale Study of Misophonia*, 74 J. CLINICAL PSYCH. 453, 453–79 (2018).

constitute freedom-of-thought infringements. Rather, the point is that some expressive acts appear to be able to violate freedom of thought. This suggests the need to consider appropriate limits on expressive action, in the interaction between speech and thought, in order to protect freedom of thought from more problematic forms of expression that speakers may employ.

V. HATE-SPEECH VIOLATIONS OF FREEDOM OF THOUGHT: CASES AND EXAMPLES

I turn now to consider specific examples of hate speech that approach or surpass a reasonable threshold of freedom-of-thought violation in the category of thought-modification. For analytical purposes, the following examples involve speech acts that combine varying levels of disturbingness, importunacy, and insidiousness. The examples vary also in terms of the kind of hate speech they include and the manner in which the hate speech is expressed. I present examples from different political and legal contexts in order to help establish the broad applicability of the framework that I provide.

A. Tabloid Treatments of the Murder of Ingrid Escamilla

Consider first an example drawn from *la nota roja* ("the red news"), a variant of "yellow journalism"[71] found in various parts of Latin America. *La nota roja* are sensationalistic tabloids that publish gruesome stories embellished with shocking images. For decades, the Mexican *nota roja* have put in plain view graphic images of murder victims, accident casualties, dead and dismembered bodies, and so forth, using shock value to draw attention and to sell newspapers.[72] In February of 2020, *¡Pásala!* and *La Prensa*, two daily Mexican tabloids with wide circulation,[73] ran stories that prominently featured leaked photographs of the body of Ingrid Esca-

---

[71] *See generally* W. JOSEPH CAMPBELL, YELLOW JOURNALISM: PUNCTURING THE MYTHS, DEFINING THE LEGACIES (Praeger Publishers eds., 2001).

[72] *See* Daniel C. Hallin, *La Nota Roja: Popular Journalism and the Transition to Democracy in Mexico*, *in* TABLOID TALES: GLOBAL DEBATES OVER MEDIA STANDARDS 267, 267–71 (Colin Sparks et al. eds., 2000). *See also* DIANA ALEJANDRA & SILVA LONDOÑO, JÓVENES EN LA NOTA ROJA: FOTOGRAFÍAS DEL HOMICIDO DE JÓVENES EN LA PRENSA DE VERACRUZ 1–26 (Athenea Digital eds., 2019).

[73] As of May 2019, *La Prensa*'s reported average daily circulation was ~287,000; for *¡Pásala!* it was 46,759. *Average daily circulation of selected paid daily newspapers in Mexico as of October 2021*, STATISTA, https://www.statista.com/statistics/1008368/newspapers-circulation-mexico (last visited Mar. 5, 2022).

milla, a 25-year-old woman who was "stabbed, skinned and dismembered" by her partner.[74] The tabloids displayed gruesome front-page, full-color images of Escamilla's corpse, taken at the crime scene, showing her mutilated body with skin and organs removed. These images were put out on display in newsstands on February 10, 2020, the day after Escamilla was murdered.[75] The headline of *¡Pásala!* read "LA CULPA LA TUVO CUPIDO" ("IT WAS CUPID'S FAULT").[76]

The tabloids' handling of Escamilla's murder prompted a backlash against the newspapers, with the entire *nota roja* genre coming under fire. Critics expressed outrage over the newspapers' treatment of the dead as well as their salacious and insensitive approach to femicide and violence against women.[77] Some critics proposed that Escamilla had been victimized twice: first by the horrific slaying and second by the lack of dignity given to her as a victim.[78] In response to Escamilla's treatment by the tabloids, Mexican feminist groups did not demand "the censorship of crimes," but instead called upon newspapers to handle gender-based violence "with more sensitivity" and to "use headlines that are less sensationalist."[79] Maria Flores, a member of Mexico's *Crianza Feminista* collective, called for "journalism with fewer adjectives," and "a straight retelling of events."[80]

The newspapers' expressive acts in this incident violated Mexican citizens' freedom of thought. Consider first the element of the disturbingness of the speech at issue, combined with the manner in which the expressive acts were presented. The crime-scene images of Escamilla's skinned and cut-up corpse, put on plain view in Mexico's public spaces, were profoundly disturbing to many who encountered them. As to the manner in

---

[74] Suman Naishadham, *Profiting from Exploitation and Violence against Mexican Women*, AL JAZEERA (Mar. 6, 2020), https://www.aljazeera.com/ajimpact/profiting-exploitation-violence-mexican-women-200305234950724.html.

[75] Daina Beth Solomon & Josue Gonzalez, *Ingrid Escamilla: Hundreds Protest Rise in Femicide in Mexico after Woman's Brutal Killing*, INDEPENDENT (Feb. 15, 2020), https://www.independent.co.uk/news/world/americas/ingrid-escamilla-killing-femicide-protest-la-prensa-mexico-a9337191.html.

[76] *See Feminicidio de Ingrid Escamilla: la indignación en México por el brutal asesinato de la joven y la difusión de las fotos de su cadaver*, BBC (Feb. 11, 2020), https://www.bbc.com/mundo/noticias-america-latina-51469528.

[77] Daina Beth Solomon, *Mexico City Killing Sparks Fury over Violence against Women*, REUTERS (Feb. 12, 2020), https://www.reuters.com/article/us-mexico-violence-gender/mexico-city-killing-sparks-fury-over-violence-against-women-idUSKBN2060CY. *Cf.* Maureen Meyer, *Combatting Femicide in Mexico: Achievements and Ongoing Challenges*, *in* THE COURAGE TO FIGHT VIOLENCE AGAINST WOMEN: PSYCHOANALYTIC AND MULTIDISCIPLINARY PERSPECTIVES 133, 133–34 passim (Paula L. Ellman & Nancy R. Goodman eds., 2017).

[78] *See* Naishadham, *supra* note 74.

[79] *Id.*

[80] *Id.*

which the speech was presented, many people exposed to the tabloid images were shocked by what they saw.[81] A significant number of people neither asked nor agreed to see the photographs, nor were they expecting to see them, nor did they wish to be exposed to the photographs or want to see Escamilla's mutilated body.[82] To the contrary, it is clear that many wished not to see the crime-scene photographs and had no intention of buying or even perusing the respective tabloids. The speech acts were therefore both deeply disturbing and expressed in importunate ways. The combination of these two factors is sufficiently weighty to create a plausible freedom-of-thought violation in the category of thought-modification.[83]

Second, the two newspapers did not just make their issues available in a single public location. Rather, the papers were displayed at newsstands in numerous public and highly visible locations across Mexico, and in quasi-public spaces such as airports, bus stations, bookstores, and so on. This made the expressive acts harder to avoid and more difficult not to experience, with many people repeatedly seeing the tabloids' respective front pages in a variety of locations. This added importunacy puts more weight on the balance in favor of a freedom-of-thought violation.[84]

Third, the tabloids' expressive acts had a significant insidiousness component regarding both femicide and murder more generally. The insidiousness of the speech did not consist merely of the salacious headlines or the stories' lewd aspects regarding intimate violence. The expressive acts contributed, if in a small way, to the desensitization of people to horrific murders and killings. This is at least a concern that weighs in favor of a freedom-of-thought violation, if not an element that is itself sufficient to count as an outright infringement, because such speech acts alter the ways in which people think and feel about terrible crimes and the victims who suffer them. To be clear, it is not, under normal circumstances, rightly a matter of governmental concern whether one desensitizes oneself to

---

[81] *Id. See also* Solomon & Gonzalez, *supra* note 75.

[82] *See* Priscilla Madrid Valero, *In Mexico, Women Are Hated to Death*, AL JAZEERA (Feb. 27, 2020), http://www.aljazeera.com/opinions/2020/2/27/in-mexico-women-are-hated-to-death.

[83] I treat each tabloid's presentation of their newspapers, on the day in question, as a single expressive act. One could call each particular exhibition of a newspaper a discrete expressive act, such that each tabloid would have performed multiple speech acts; that would not affect the argument at hand.

[84] Exposure to single disturbing speech acts will in many cases not rise to the level of freedom-of-thought infringement. *Cf.* Cohen v. California, 403 U.S. 15 (1971) (upholding the legal permissibility of public displays of expletives as emotive speech, where the speech is neither an incitement to violence nor unable to be avoided by others). The framework I provide allows that when expressive acts are not only disturbing but also importunate, they move nearer to violating freedom of thought, all else being held equal.

femicide or to murder more generally.[85] However, repeatedly to shock others and to dull their sense of the horror and outrage of such crimes can plausibly be not just a moral problem but also an act warranting government involvement. This point holds even if the greater proportion of a population no longer finds the expressive acts at issue to be particularly disturbing; and there are always young people and not-yet-desensitized adults whose positions must be considered.

Whether any one of the particular elements of disturbingness, importunacy, or insidiousness is adequate to establish a thought-modification violation of freedom of thought, in this example, the combination of factors is sufficient. The example appears clearly to involve hate speech on the part of the two tabloids, as well. The callous and exploitative treatment of Escamilla, and of femicide, communicated hatred of women and it undermined women's status as free and equal citizens. The point is robust whether the tabloids' editors and publishers had any intense dislike of Escamilla or hatred of women in general. As I have suggested, hate speech can exist even when the speakers may have no hate for the parties to whom their expression applies, and whether they are motivated by intense or passionate dislike when supplying the expression to which they give voice.

The argument that I provide suggests that the tabloids would violate people's freedom of thought even by showing anonymous murder victims on front page, ceteris paribus, instead of depicting women, named individuals, or both. But that does not detract from two key points: that hate speech can violate freedom of thought, and that hate speech toward women is a key element of the expressive act that infringes freedom of thought, in this example.

B. Slave-Auction Markers in Virginia and Maryland

Some American localities incorporate, in public spaces, deeply disturbing and insidious markers of slavery in America. Examples include the slave-auction block in downtown Fredericksburg, Virginia (which was removed on June 5, 2020),[86] a purported "Old Slave Block" on a public street-corner in Sharpsburg, Maryland,[87] a purported "Old Slave Block" on a public street-corner in Sharpsburg, Maryland,[88] and a public sidewalk

---

[85] *See generally* SWAINE, *supra* note 17, for the proposition that it is in many cases morally wrong to desensitize oneself to such crimes.

[86] *See Slave Auction Block*, FXBG, https://www.fredericksburgva.gov/1287/Slave-Auction-Block (last visited Mar. 5, 2022); *Slave Auction Block Relocation*, FXBG, https://www.fredericksburgva.gov/1680/Slave-Auction-Block-Relocation (last visited Mar. 5, 2022). The Fredericksburg slave-auction block has been relocated to the Fredericksburg Area Museum.

[87] *See* Alexis Fitzpatrick, *Sharpsburg Removes "Slave Auction Block" for Restoration*, HERALD-MAIL MEDIA (June 19, 2020), https://www.herald-mailmedia.com/story/news/2020/06/19/sharpsburg-removes-slave-auction-block-for-restoration/43670765/.

[88] *Id.*

plaque marking the site of slave auctions in the Court Square area of downtown Charlottesville, Virginia.[89] While the markers are not importunate expressive acts, given that they are fixed and individual, and appreciating that they can largely be avoided, they are examples of deeply disturbing and highly insidious racial hate speech. They are disturbing in the ways in which they memorialize the slave trade, commemorating places where people actually stood to be sold as slaves. Importantly, such markers have been placed or left in public locations, conspicuously lacking explanation, contextualization, or commentary regarding the terrible suffering and injustice done to enslaved people. They give no sense even of the basic wrongness of the slave trade: in each of the cases mentioned above, the respective markers refer to "slaves" in their brief identificatory text and not to "people" or "human beings."[90]

The markers are also highly insidious expressive acts. This becomes clear when one considers: (a) what the language they employ communicates, (b) what is transmitted by what is missing and unstated at the markers, (c) their public placement, and (d) the actions that have not been taken, at the sites in question, to remedy problems with what the memorial sites convey. It is also telling that public officials failed to include, at the respective sites, even brief commentary on the wrongness or the trauma caused by the slave trade.[91] These factors are compounded by the normality of the stated descriptions given at each location. The plaque on the so-called "Old Slave Block" in Sharpsburg, for instance, affirms that the stone is a "famous landmark" and states that it has been so for over 150 years.[92] Similarly, the sidewalk plaque in downtown Charlottesville read: "SLAVE AUCTION BLOCK On this site slaves were bought and sold."[93] When the plaque was stolen in early 2020, Jalane Schmidt reflected on the "disturbing" nature of the plaque's removal, remarking that while "the slave auction plaque was so small and set in the ground [that] you could

---

[89] *See* Neal Augenstein, *Guilty Plea for Man Who Admitted Taking Charlottesville Slave Auction Marker*, WTOP NEWS (June 5, 2020), https://wtop.com/virginia/2020/06/guilty-plea-for-man-who-admitted-taking-charlottesville-slave-auction-marker.

[90] *See Fredericksburg Slave Auction Block Site*, CLIO, https://www.theclio.com/entry/7178 (last visited Mar. 5, 2022); *Old Slave Block*, HIST. MARKER DATABASE, https://www.hmdb.org/m.asp?m=185589 (last visited March 5, 2022); Augenstein, *supra* note 89.

[91] The City of Fredericksburg has taken steps to try to "[honor] all voices in history at the former site of the Slave Auction Block." Measures include installing "[a] new wayside panel display" at the site in question and developing a "permanent interpretation." *See City of Fredericksburg Honors the Importance of a More Inclusive History at the Former Site of the Slave Auction Block*, FXBG (Oct. 27, 2020), http://www.fredericksburgva.gov/DocumentCenter/View/18200/News-City-Honors-Importance-of-More-Inclusive-History-102720.

[92] The "Old Slave Block" was vandalized in early June of 2020 and the stone has been removed "for cleaning and restoration." *See* Fitzpatrick, *supra* note 87.

[93] *See* Augenstein, *supra* note 89.

walk over it, it was the only thing we had to commemorate the slaves whose lives were torn apart there."[94]

The slave-auction plaque might have been the only thing placed in memory of enslaved people in Charlottesville's Court Square, but it is worth considering how or whether that may be problematic, especially given the amount of Confederate memorialization in the area. In addition, while it is true that one could walk over the Charlottesville slave-auction plaque, the plaque was in a sense made to be stepped on, and people have trodden upon it many times over the years. These slave-auction markers are noxious and artful reminders of slavery and inequality in America, conveying historical and ongoing disrespect for African Americans. Those who must live with such markers experience messages that subtly normalize, if not venerate, the inferiority of African Americans.[95] There is no pressing reason or special educational purpose to having the slave-auction markers remain as they are, in America's public spaces, and for people to continue to encounter their disturbing and insidious messages.

The slave-auction markers satisfy the criteria for hate speech that I have described: they memorialize and communicate disrespect and contempt for African Americans. The harms are sufficiently serious, in terms of how disturbing and insidious they are, and they can reasonably be seen as violations of freedom of thought that call for governmental action. There are various ways in which government officials might proceed to remedy the problems at hand, depending on relevant jurisdictional powers and existing rules and laws. The value of freedom of thought provides no determinate or specified remedies for such situations, although it is action-guiding with respect to measures to take. The respective officials could, for instance, cover, move, or permanently remove the markers and plaques. Another possibility would be to curate the sites more comprehensively, adding or amending language and providing greater context. Officials could also replace the existing markers and plaques with new memorials that are neither deeply disturbing nor highly insidious. And it is possible for public markers and monuments to express proper and appropriate public disapproval of such practices as slave auctions. Steps like these would not be overly costly or burdensome, and they appear to be warranted for monuments and plaques of the kind I have described.

---

[94] Michael E. Miller, *A Stolen Slave Auction Plaque Shook Charlottesville. But the Confession Was the Real Shock*, WASH. POST (Feb. 11, 2020), https://www.washingtonpost.com/history/2020/02/10/charlottesville-slave-plaque-stolen. *See also* Augenstein, *supra* note 89. The City of Charlottesville has not yet replaced the plaque, which was returned, in the sidewalk at Court Square.

[95] *Cf.* DELGADO & STEFANCIC, *supra* note 4, at 32, 143–44 (explaining that Critical Race Theory affirms that hate speech "teaches minorities 'their place'").

C. R.A.V. v. City of St. Paul

R.A.V. v. City of St. Paul provides another example of hate speech that violates freedom of thought.[96] That case concerned a juvenile who burned a makeshift medium-sized cross on the front lawn of an African American family in St. Paul, Minnesota, on June 21, 1990.[97] St. Paul had in place a Bias-Motivated Crime Ordinance prohibiting displays of symbols that a person knows, or has reason to know, "arouses anger, alarm or resentment in others on the basis of race, color, creed, religion or gender"; this included such symbols as "a burning cross or Nazi swastika."[98] St. Paul's ordinance was supposed to cover only speech that amounted to "fighting words," but the Supreme Court determined that the statute wrongly prohibited speech solely on the basis of subjects it addressed.[99] The justices held unanimously that the Minnesota law was unconstitutional. Justice Antonin Scalia, writing for the Court, stated that government may not "impose special prohibitions on … speakers who express views on disfavored subjects."[100] He suggested that Minnesota's lawmakers could have instead employed content-neutral alternatives in their legislation to produce "the same beneficial effect" as the one sought by the ordinance.[101] Concurring justices voiced concern that the statute was overbroad, criminalizing a considerable amount of speech protected by the First Amendment.[102]

Whatever one might think of the constitutionality or the normative soundness of Minnesota's Bias-Motivated Crime Ordinance, it is plausible that the expressive act that kicked off *R.A.V.* sufficed as a violation of freedom of thought. However, for analytical clarity, it serves to consider a slightly stylized example in which an individual ignites a large cross in a public space directly in front of another person's house, departing as soon as the cross is lit.[103] The action in this example could knowably cause, and

---

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

[97] *Id.* at 379.

[98] *Id.* at 380.

[99] *Id.* at 380–81. *Cf.* Chaplinsky v. New Hampshire, 315 U.S. 568, 571–72 (1942) (proposing that prevention and punishment of fighting words, along with speech that is "lewd and obscene," "profane," or "libelous," has "never been thought to raise any Constitutional problem").

[100] *R.A.V.*, 505 U.S. at 391.

[101] *R.A.V.*, 505 U.S. at 395–96.

[102] *R.A.V.*, 505 U.S. at 397 (White, J., concurring).

[103] This example is different from the *R.A.V.* case inasmuch as the symbol in question (i.e., the burning cross) is considerably larger and the cross-burning is not performed, without permission, in someone's front yard. Cross-burnings executed at private residences, without the respective owners' permission, elicit added issues of trespass, arson, criminal damage to property, and so forth, *see R.A.V.*, 505 U.S. at 380 nn.1–2; *see also* STROSSEN, *supra* note 1, at 54–56. The complicating factors mentioned here also raise "fighting words" considerations. *See Chaplinsky*, 315 U.S. at 573; Terminiello v. City of Chicago, 337 U.S. 1 (1949); Feiner v. New York, 340 U.S. 315 (1951); Texas v. Johnson, 491 U.S. 397

be known to be highly likely to cause, "anger, alarm, or resentment in others on the basis of race, color, creed, religion, or gender[.]"[104] The burning cross is a proud and frightful symbol of the Ku Klux Klan, used for years to terrorize the Klan's enemies.[105] The problem is that it is very easy to arouse anger, alarm, or resentment in people, on any of the listed bases or otherwise. One would need only to generate knowably any one of the reactions, furthermore, in order to satisfy the stated criterion.

In addition, a civil or criminal ordinance based on such a standard would fail to consider whether the affected parties were reasonably angered, or alarmed, or made resentful. It would suffice simply that they were provoked in the ways described. In his concurring opinion in *R.A.V.*, Justice Byron White reflected that the "mere fact that expressive activity causes hurt feelings, offense, or resentment does not render the expression unprotected" under American constitutional law.[106] The Minnesota ordinance does not refer expressly to hurt feelings,[107] but it must be acknowledged that: (a) some people's feelings are very easy to hurt, (b) some people get hurt feelings for practically no reason, and (c) people's feelings might at times be hurt unavoidably, in otherwise entirely appropriate and lawful speech or conduct. Hurt feelings, like offense,[108] are not a sound basis on which to limit speech, much less to criminalize it.

Justice Harry Blackmun regretted that the Court's decision in *R.A.V.* would, if it were to set precedent, prevent states from "regulat[ing] speech that causes great harm unless [those states] also regulate[d] speech that does not."[109] Blackmun proposed:

> I see no First Amendment values that are compromised by a law that prohibits hoodlums from driving minorities out of their homes by burning crosses on their lawns, but I see great harm in preventing the people of Saint Paul from specifically punishing the race-based fighting words that so prejudice their community.[110]

Many might agree with Justice Blackmun's suggestion that the First Amendment is not at odds with laws protecting people from having crosses burned on their lawns. But the "fighting words" doctrine seems

---

(1989); Virginia v. Black, 538 U.S. The stylized example obviates the aforementioned complications for analytical purposes.

[104] *R.A.V.*, 505 U.S. at 380.

[105] *See* WYN CRAIG WADE, THE FIERY CROSS: THE KU KLUX KLAN IN AMERICA 144, 224, 276, 419 (1987).

[106] *R.A.V.*, 505 U.S. at 414 (White, J., concurring).

[107] *See* St. Paul Bias-Motivated Crime Ordinance, St. Paul, Minn., Legis. Code §292.02 (1990) (cited in *R.A.V.*, 505 U.S. at 380).

[108] *See* COHEN, *supra* note 68, at 124, 131.

[109] *R.A.V.*, 505 U.S. at 415 (Blackmun, J., concurring).

[110] *Id.* at 416 (Blackmun, J., concurring).

inadequate to handle the stylized cross-burning case, given that the incendiarist does not remain to taunt or to incite the residents. For that matter, Justice Blackmun does not explain what the great harm is in the speech at issue in *R.A.V.*, although his words are suggestive in that respect.

Freedom of thought considerations better account for the issues and concerns that the stylized case raises, drawing together key factors in a more coherent way. The concepts and categories I have outlined provide a basis on which to argue that, in the stylized example, the expressive act of cross-burning is sufficiently disturbing, importunate, and insidious to violate freedom of thought. The framework allows one more properly to account the for the harms in the stylized case, and it gives a fuller sense of their gravity, as well. To take each axis in turn: first, the cross-burning is deeply disturbing inasmuch as it provides a frightening and lasting impression. It is profoundly shocking, conveying the horror of racial violence and communicating extreme hatred of members of racial and ethnic minorities. Not only does the act convey the message that the targeted people had better get out, it communicates that they—like other victims in the past—should fear their own murder, punctuating the statement with the symbolic violence of flames. Second, the expressive act is highly importunate: one cannot readily or easily avoid a large burning cross in front of one's house. It is a spectacle that one can hardly ignore, with the sight, sound, smell, and heat of the flaming cross affecting most if not all of one's senses. Third, the burning cross is a very insidious expressive act. It transmits, along with its loud and brazen messages, subtler connotations of lawlessness and extralegal violence, of the victims' inferiority and their lack of equal protection under the law.

VI. Freedom of Thought and Hate-Speech Regulation

Having outlined and discussed specific examples of hate-speech violations of freedom of thought, I move now to consider what the present analysis offers with respect to hate-speech regulation. In particular, I consider what this account may provide on five salient issues: (a) regulating speech based on content or viewpoint, (b) whether hate speech should be treated as "fighting words," (c) the compatibility of the present argument with other theories of hate speech, (d) how the argument may fit with different approaches to hate-speech regulation, and (e) how the case that I present fares with respect to notable objections to hate-speech regulation. The points that I supply may be calibrated for treatments of hate speech and freedom-of-thought infringements in different political and legal contexts.

First, allowing government to regulate hate speech on the basis of freedom-of-thought concerns would neither require nor entail content-based speech regulation or viewpoint discrimination. Expressive acts

could be assessed on viewpoint- and content-independent metrics,[111] with no limits placed on people's viewpoints and no restrictions applied to content of expression per se. The working conception of hate speech facilitates this prospect by including content-independent criteria that expressive acts would need to satisfy in order to count as hate speech. Nor does this approach give cause to expect any particular kind of content or viewpoint to be regulated more than others.[112] Hate speech provides good examples of thought-modification violations, but that does not necessarily militate in favor of stricter or greater regulation of those expressive acts, in comparison to forms of expression that are not hate speech, ceteris paribus. What is more, the argument that I provide allows that particular instances of hate speech need not necessarily violate freedom of thought. Subjecting expressive acts to regulation if they violate freedom of thought does not require restrictions on content or on viewpoint, and regulation based on the elements I have discussed would be consistent with permitting speakers freely to express and to circulate their ideas using means of expression that are not deeply disturbing, insidious, or highly importunate.

Note also that the framework that I describe allows for the possibility that any speech act that is sufficiently disturbing, importunate, or insidious could be a candidate for regulation.[113] Even a reading of the Declaration of Independence could infringe freedom of thought if the expressive act were highly importunate (e.g., performed repeatedly and at high volume, in front of someone's residence), disturbing (e.g., including particularly gruesome imagery), or insidious (e.g., using untoward and excessive subliminal techniques). In addition, the very idea of freedom of thought helps to make sense of why government is charged with the responsibility of regulating annoyances that have no expressive content. This includes the regulation of nuisances and affrays, ordinances covering "noisy sound truck[s],"[114] bylaws regulating annoying sights and smells, and so on. In

---

[111] *See* Packingham v. North Carolina, 137 S. Ct. 1730 (2017); Reed v. Town of Gilbert, 576 U.S. 155 (2015); Rosenberger v. Rectors and Visitors of the University of Virginia, 515 U.S. 819 (1995). *Cf.* Walker v. Texas Division, Sons of Confederate Veterans, 576 U.S. 200 (2015) (determining that viewpoint discrimination does not apply to government speech).

[112] *See* Barendt, *supra* note 4, at 543 (discussing minority-group members expressing hate speech toward members of majority groups); *see also* ERIC BARENDT, FREEDOM OF SPEECH 178 (2d ed. 2005).

[113] Mere exposure to unwanted or unagreed-to speech acts is not itself sufficient for freedom-of-thought infringement. Reasonable communicative and free-speech standards require the allowance of such exposure. One could not reasonably assert a freedom-of-thought right against exposure to such speech unless other criteria warranting governmental involvement were present. *Cf.* Fed. Commc'ns Comm'n v. Pacifica Found., 438 U.S. 726 (1978).

[114] *R.A.V.*, 505 U.S. at 386. Both a "sound truck" and fighting words "can be used to convey an idea," explains Justice Scalia, "but neither has, in and of itself, a claim upon the First Amendment." *Id.* Trucks' sounds in many cases do not constitute expressive acts. Not only are those noises not considered protected

short, unpopular viewpoints would not be bases for regulation, no particular content would be proscribed, regulation would not be limited to the "disfavored topics" of race, color, creed, religion, or gender,[115] and the framework would not be unduly prone to regulating speech on some topics but not others. This would allow appropriate and reasonable regulation to pass intermediate scrutiny, at least on the American example; the framework would "leave open ample alternative channels for communication of […] information,"[116] and it could be companionable with similar commitments in other countries.

Second, the argument that I have marshaled does not construe hate speech as fighting words, at a conceptual level, nor does it support the surmise that the doctrine of fighting words adequately covers freedom-of-thought concerns. The fighting-words doctrine emerged out of the United States Supreme Court case *Chaplinsky v. New Hampshire*, wherein such language was characterized as words that "by their very utterance inflict injury or tend to incite an immediate breach of the peace."[117] Fighting words were in *Chaplinsky* given a low level of protection under law, on the rationale that such utterances "are no essential part of any exposition of ideas."[118] Fighting words have only "slight social value," the Court proposed, such that "any benefit that may be derived from them is clearly outweighed by the social interest in order and morality."[119] The scope of fighting words has since been narrowed: in *Terminiello v. City of Chicago*, the Court determined that speech that merely causes unrest is not enough to count as fighting words.[120] The expression must "[be] shown likely to produce a clear and present danger" well above mere "public inconvenience, annoyance, or unrest" in order to be restricted.[121] The Court's current standard is even tighter, specifying that fighting words constitute a "small class" of speech and proposing that those words consist of "direct

---

speech under the First Amendment, the value of freedom of thought seems to justify, and perhaps also implicitly motivates, their regulation.

[115] *R.A.V.*, 505 U.S. at 391; *see also id.*, 395–96, 404; *cf.* STROSSEN, *supra* note 1, at 94–99 (providing examples of regulation of disfavored views).

[116] *See* Clark v. Cmty. for Creative Non-Violence, 468 U.S. 288, 293 (1984); *see also* Ward v. Rock Against Racism, 491 U.S. 781, 783, 791, 802–03 (1989).

[117] *Chaplinsky*, 315 U.S. at 572.

[118] *Id.*

[119] *Id. Cf.* ZECHARIAH CHAFEE, JR., FREE SPEECH IN THE UNITED STATES 149–52 (1941).

[120] Terminiello v. City of Chicago, 337 U.S. 1 (1949).

[121] *Id.* at 4; *cf.* Laura Beth Nielsen, *Power in Public: Reactions, Responses, and Resistance to Offensive Public Speech*, *in* SPEECH & HARM: CONTROVERSIES OVER FREE SPEECH 148, 171 (Ishani Maitra & Mary Kate McGowan eds., 2012) (arguing that annoying or offensive speech receives different treatment depending on whether the speech targets "more privileged members of society," referring to begging as an example). *Contra Terminiello*, 337 U.S. at 26 (Jackson, J., dissenting).

personal insult or an invitation to exchange fisticuffs."[122] This has meant that flag-burning does not constitute fighting words, with the expressive act of flag-burning counting as protected speech under the First Amendment.

These constitutional developments indicate that the fighting-words doctrine is too narrow to cover the concerns about hate speech and freedom of thought that I have canvassed. The central problem is that hate speech can violate freedom of thought without constituting a provocation to fight or an insult of any obvious kind, personal or otherwise. Hate speech can insult or provoke in ways that satisfy the narrow delineation of fighting words in *Texas v. Johnson*,[123] and hate speech that infringes freedom of thought can be provocative or insulting in the relevant ways, as well. But hate speech can infringe freedom of thought without doing any direct personal affrontery, just as it need not arouse ire in those subjected to it—the tabloid and slave-auction marker cases, discussed above, are examples. These factors illuminate how the fighting-words doctrine neither suffices to cover hate-speech concerns nor adequately handles the various forms of disturbingness, importunacy, or insidiousness found in speech-based violations of freedom of thought. The analytical framework that I provide has the advantage of not linking hate speech and fighting words so tightly.

Third, the argument that I present is compatible with other theories of the harms in hate speech, and it may potentially complement other justifications for hate-speech regulation, as well. First of all, the harm inherent in hate-speech-based infringements of freedom of thought does not compete with other commonly identified harms in hate speech. The list of harms associated with hate speech includes, as I have noted, claims to the effect that hate speech perpetuates subordination, incites violence, risks physical harm, constitutes or causes psychic wounding, defames people, impugns people's character, is libelous, and so forth. The harm that I identify is a distinct rights-violation that is separable analytically from the above listed harms. It is addable to those harms, and it is compatible with them inasmuch as particular instances of hate speech may constitute or produce both freedom-of-thought violations as well as other harms on the list. What is more, the harm of a freedom-of-thought encroachment can also be integrated with other working accounts of hate-speech harms, and so it is in that sense complementary at a theoretical level.

The finding that hate speech can infringe freedom of thought also contributes to the justification of regulation of hate speech by political and legal institutions. I have argued that hate speech violates a crucial right that people should be acknowledged to have and which government ought to protect. There are various ways in which the concern to protect freedom

---

[122] Texas v. Johnson, 491 U.S. 397, 409 (1989); *see also* Snyder v. Phelps, 562 U.S. 443 (2011).

[123] *See generally Texas*, 491 U.S.

of thought can add to justificatory frameworks favoring hate-speech regulation. For example, the considerations mentioned here might complement or support other accounts aiming to justify restrictions on hate speech. Another possibility is that freedom-of-thought considerations provide a more solid normative foundation on which to justify government regulation of certain manifestations of hate speech, beyond what other accounts provide. The basis for regulating hate speech that I identify adds justificatory weight in favor of regulating the segment of hate speech that constitutes or causes thought-modification violations, and it can contribute to positions favoring hate-speech regulation, by so doing.

Fourth, the present argument allows that there are many possible ways in which liberal democracies might reasonably proceed, with respect to devising and implementing legal structures for protecting freedom of thought from hate speech. In terms of differential legal classifications and distinctions between criminal and civil law, the framework that I propose recommends treating many hate-speech infringements as civil offenses. It is in this respect unlike the Minnesota ordinance at issue in *R.A.V.*[124] However, the schema does not exclude the possibility of criminal penalties for hate-speech encroachments on freedom of thought, and it can be fitted with approaches to regulation that include both civil and criminal forms of penalization.[125] It is sensible to allow that some freedom-of-thought violations ought to be treated as criminal offenses. Similarly, one could reasonably conceive of some other speech-related crimes, such as certain forms of intimidation, as freedom-of-thought infringements.[126] In addition, the present argument affirms the appropriateness of criminalizing infringements of freedom of thought in graver cases in which expressive action may not be involved, such as in those associated with wrongful medical manipulation or experimentation of the sort mentioned previously.

The framework for protecting freedom of thought that I have discussed also recommends time, place, and manner restrictions on speech acts.[127] The three thought-modification axes are auspicious for devising an enhanced and more thorough framework for such regulation. Those

---

[124] *See R.A.V.*, 505 U.S. 377; *see also Snyder*, 562 U.S.

[125] *See* Maxime Lepoutre, *Hate Speech Laws: Expressive Power Is Not the Answer*, 25 LEGAL THEORY 272, 273–74, 283, 290 (2019).

[126] *See* Virginia v. Black, 538 U.S. 343, 343–45, 354–75, 383–84; *see also* Watts v. United States, 394 U.S. 705 (1969).

[127] *See* Clark v. Cmty. For Creative Non-Violence, 468 U.S. 288, 293 (1984); Ward v. Rock Against Racism, 491 U.S. 781, 781–82, 798–800 (1989); R.A.V. v. City of St. Paul, 505 U.S. 377, 386 (1992); Heffron v. Int'l Soc'y for Krishna Consciousness, 452 U.S. 640 (1981); Packingham v. North Carolina, 137 S. Ct. 1730, 1735, 1739 (2017); Hague v. Comm. For Industrial Org., 307 U.S. 496, 515 (1939). *Cf.* Kevin McGravey, *Reimagining the First Amendment: The Assembly Clause as a Substantive Right*, 53 FIRST AMEND. STUD. 67, 72–74 (2019); TIMOTHY ZICK, SPEECH OUT OF DOORS: PRESERVING FIRST AMENDMENT LIBERTIES IN PUBLIC PLACES 54, 57–61, 117–19, 204, 244, 277–82 (2008).

axes help to clarify what is important about freedom of thought and they offer a fuller sense of how speech-acts can put freedom of thought in jeopardy. The justificatory reasons for shielding freedom of thought from excessively disturbing, importunate, or insidious speech assist in providing new and better grounds for time, place, and manner restrictions, as well. A reworked regulatory structure would have the added benefit of being able to cover a full range of expressive acts, hate speech included, without singling out particular topics, viewpoints, or categories of speech for special treatment. This gives further reason to hold that the framework I furnish would not easily be used to silence minority groups' expression, nor does it raise standard "slippery slope" concerns about hate-speech regulations opening the door to pervasive legal suppression of speech.

Fifth, the regulation of expressive acts according to reasonable freedom-of-thought concerns would permit broad expression of ideas without restricting opinions or views.[128] Not only is the position that I advance entirely compatible with rejection of prior restraint on speech,[129] but a wide range of content, viewpoints, and opinions would be able freely to circulate throughout society, on a full range of topics and issues, subject to reasonable restrictions of the kinds I have distinguished. These advancements would provide a structure in which citizens could engage in speech and counterspeech, without relying upon counterspeech to handle freedom-of-thought violations.[130] This approach would help to establish people's reasonable expectations regarding the different kinds of expressive acts they may encounter as they move between spheres in public and private life. It would also mitigate longstanding concerns about hate-speech regulations forcing detestable views underground. And deft protection of freedom of thought from hate speech would support democratic norms favoring criticism of such social ills as racism or sexism, addressing an important concern of Edwin Baker's.[131] The kinds of regulation that I describe could facilitate frank and open discussion and facilitate growth of democratic self-understanding, strengthening and advancing norms of civility.[132]

The value of freedom of thought does not simply countenance or begrudge people their expressive liberty or their freedom of opinion—it endorses and bolsters those freedoms. A fuller recognition and integration of

---

[128] *Cf.* STROSSEN, *supra* note 1, at 13–14, 27–29, 69–104, 105–19; Post, *supra* note 11, at 123, 123–25; Baker, *supra* note 3, at 141, 142, 156.

[129] *See* Near v. Minnesota, 283 U.S. 697 (1931); Reno v. Am. Civ. Liberties Union, 521 U.S. 844 (1997); *Hague*, 307 U.S. at 516; Cantwell v. Connecticut, 310 U.S. 296, 307 (1940); *cf.* Fed. Commc'ns Comm'n v. Pacifica Found., 438 U.S. 726, 726 (1978).

[130] *See generally* Lepoutre, *supra* note 125; *see also* Paul Billingham, *State Speech as a Response to Hate Speech: Assessing 'Transformative Liberalism'*, 22 ETHICAL THEORY & MORAL PRAC. 639 (2019).

[131] Baker, *supra* note 3, at 151.

[132] *See id.* at 152–53.

freedom of thought in liberal democracies could help to enhance discussion and to create conditions sustaining better forms of discursive interaction. It would allow opinions and views to be heard and assessed, and criticized, and it would support the robust protection of one's right to dissent. Taken together, these factors illuminate how a well-formed schema for protecting freedom of thought from deleterious and harmful speech acts would not drive ideas and viewpoints out of the marketplace of ideas,[133] nor would it suppress the practices, or the freedoms, of individual thought, speech, or opinion.

CONCLUSION

I have argued in this Article that hate speech can infringe freedom of thought. The account that I provide allows one to describe harms in hate speech that go beyond mere hurt feelings, disquietude, or simple offense, and which are different from fighting words, incitement, libel, social unrest, and related harms. My account also gives clearer expression to significant harms that are constituted or caused by hate speech. And it has the benefit not just of fitting with widespread concerns for combating hate speech but of applying to other forms of expression than hate speech, and to incursions on freedom of thought that are not speech-based, increasing the power and the applicability of the overall approach.

The theoretical structure for hate-speech regulation that I outline has the advantage of being compatible with other theories of hate speech and with different views of what the harms in hate speech happen to be.[134] The case that I furnish may prove companionable with a variety of standpoints regarding freedoms of speech and of thought, furthermore. It is flexible with respect to conceptions of the nature of those freedoms, and regarding how and when those freedoms may be violated. One hopes that the arguments in this Article may contribute positively to theoretical and practical treatments of hate speech and assist in clarifying liberal democracies' central constitutional commitments.

***

---

[133] *Cf.* Simon & Schuster v. Members of the N.Y. State Crime Victims Bd., 502 U.S. 105 (1991); Abrams v. United States, 250 U.S. 616, 630–31 (1919) (Holmes, J., dissenting); MILL, *supra* note 58.

[134] This includes differentiation across possible views of whether hate speech constitutes or causes harms that constitute (or cause) freedom-of-thought infringements; *see generally* Barendt, *supra* note 4.