

HATE SPEECH—DEFINITIONS & EMPIRICAL EVIDENCE

*Katharine Gelber**

James Weinstein's paper is a thoughtful, refreshing and considered contribution to the ongoing debate over whether or not hate speech laws can be justified in liberal democratic orders.

As a political scientist who has spent 20 years investigating, analyzing, and reporting on the introduction, implementation, and effects of hate speech laws, I wish to focus in my commentary on the empirical assumptions and claims that inform Weinstein's argument. I will focus on three. The first is his failure to conceptualize hate speech in a way that is commensurate with much of the philosophical literature focussing on definitions of hate speech. The second is his apparent reluctance to concede that hate speech (in both vituperative and more modest forms) is capable of harming political participation to the degree he would require, or in the ways he would require, in order for hate speech laws to be justifiable in political legitimacy terms. The third are his empirical claims about the operation of hate speech laws in practice.

WHAT IS HATE SPEECH?

On the first point, Weinstein's article lacks a clear conception of hate speech. To be fair, it is part of Weinstein's argument that the lack of a clear definition of hate speech is in fact part of the problem. In part, I agree with him. In particular, I find the use of the term 'hate' to be misleading in this context, since it implies that the defining feature of hate speech is virulent dislike of a person for any reason. As I will explain below, this is not my, or

* Professor of Politics and Public Policy, University of Queensland. I am grateful to the Australian Research Council for financial support (FT110100114, DP1096721), and to Adrienne Stone, Susan Brison, Alex Brown, and Jim Weinstein for helpful discussions on the thoughts in this article.

others', view of the defining feature of hate speech. Nevertheless, it is a term that is widely used in the literature.

Weinstein's complaint about the lack of definition of hate speech, however, is different from mine. He claims that, in spite of Waldron's claim that hate speech laws only restrict people from expressing their views in particularly vituperative ways, and not from expressing their points of view on any topic, hate speech laws do in fact result in poor applications in practice. He argues further that these applications put people through difficult and time consuming legal processes without good reason. I will deal separately with this latter claim below.

Weinstein navigates through this component of his argument by describing hate speech in his introduction as expression that "demeans."¹ Later, when discussing a case adjudicated in the European Court of Human Rights, he suggests that because Gillmerveen's discourse was "odious" but had "little vituperation and no use of epithets," it was a case of the misapplication of hate speech laws.² He similarly laments the lack of vituperation and/or epithets in some of his other examples of the misapplication of hate speech laws. He implies quite strongly, therefore, that (at least) two of the defining features of hate speech, properly understood, are vituperation and the use of epithets.

With respect, this pays no attention to the considerable body of literature that has developed over the last few decades, which discusses how hate speech is expression that materially and substantively harms its targets in the saying of that speech (and not only in terms of a discreet, consequential harm arising from it). These speech-act theory informed perspectives both on racist hate speech³ and on pornography⁴ argue that harm can occur whether the hate speech is expressed in vituperative terms or not, and whether epithets are used or not. According to this literature, the defining features of hate speech are not whether it is

1. James Weinstein, *Hate Speech Bans, Democracy and Political Legitimacy*, 32 CONST. COMMENT. 527, 527 (2017).

2. *Id.* at 554.

3. See, e.g., SPEECH AND HARM: CONTROVERSIES OVER FREE SPEECH (Ishani Maitra & Mary Kate McGowan eds., 2012); WORDS THAT WOUND: CRITICAL RACE THEORY, ASSAULTIVE SPEECH, AND THE FIRST AMENDMENT (Mari J. Matsuda et al. eds., 1993).

4. See, e.g., Rae Langton, *Speech Acts and Unspeakable Acts*, 22 PHIL. & PUB. AFF. 293, 297 (1993). See also Rae Langton, *Whose Right? Ronald Dworkin, Women, and Pornographers*, 19 PHIL. & PUB. AFF. 311, 311-59 (1990).

vituperative and whether or not it contains epithets, but that it incurs harms discursively when the hate speech is uttered, and that these harms are analogous to other discriminatory harms, such as denying someone a service or denying them a job on the ground of their race or other relevant attribute.

Indeed, in this context Waldron acknowledges that he is not the first theorist to have developed a liberal argument as to the viability of regulating hate speech. In the late 1980s and early 1990s, Ronald Dworkin engaged in a fierce debate with other philosophers on the question of the regulation of pornography. In that debate, Dworkin argued that the regulation of pornography was not justifiable on democratic grounds. He viewed the case for the harm of pornography as inconclusive, and the silencing argument⁵ (the argument that pornography as a type of hate speech operates to “silence” women by rendering their protestations against sexual mistreatment unable to achieve their intended outcome of stopping that mistreatment, and by rendering them unable to be heard as authentic and dignified speakers) as unconvincing.⁶ In response, Jennifer Hornsby argued that he misunderstood the claim that pornography silences women.⁷ Rae Langton argued, using a Dworkinian version of liberalism, that it was possible to make out a consistent case that pornography ought to be regulable as a violation of women’s civil rights.⁸ Jeremy Waldron also acknowledges, in his response to Weinstein in this volume, that Dworkin’s view of political legitimacy can sustain an argument that hate speech laws themselves can contribute to democratic legitimacy.

The debate between Weinstein’s response in this volume and Waldron’s original piece echoes that earlier one, in the sense that

5. For the most influential arguments on the silencing effects of pornography, see CATHARINE A. MACKINNON, *ONLY WORDS* (1993); Rae Langton, *Beyond Belief: Pragmatics in Hate Speech and Pornography*, in *SPEECH AND HARM: CONTROVERSIES OVER FREE SPEECH* 72–93 (Ishani Maitra & Mary Kate McGowan eds., 2012); Rae Langton, *Subordination, Silence, and Pornography’s Authority*, in *CENSORSHIP AND SILENCING: PRACTICES OF CULTURAL REGULATION* 261–84 (Robert C. Post ed., 1998); Rae Langton, *Speech Acts*, *supra* note 4.

6. Ronald Dworkin, *Two Concepts of Liberty*, in *ISAIAH BERLIN: A CELEBRATION* 100–09 (Edna Ullmann-Margalit & Avishai Margalit eds., 1991); Ronald Dworkin, *Women and Pornography*, N.Y. REV. BOOKS, Oct. 21, 1993, at 36, 37, 40–42.

7. Jennifer Hornsby, *Speech Acts and Pornography*, in *THE PROBLEM OF PORNOGRAPHY* 220–32 (Susan Dwyer ed., 1995)

8. Rae Langton, *Whose Right? Ronald Dworkin, Women, and Pornographers*, 19 *PHIL. & PUB. AFF.* 311, 311–59 (1990).

Weinstein speaks past, and appears not to hear or recognize, the claims that hate speech acts harm in the saying of them, and that they harm in important ways that silence their targets. This is not to say that every instance of hate speech must silence its target, as clearly some instances of hate speech result in wide public disapprobation, or mobilise counter claims by targets and their allies. But it is to say that hate speech properly understood is capable of harming in these ways, and that these harms ought to be recognized and interpolated into their arguments in ways that Weinstein has not done.

Understanding speech-act informed claims about how hate speech harms is important, because if Weinstein had conceived of hate speech in these terms, it would have had significant implications for his argument. It would among other things have undermined his implication that it is only by *not* regulating hate speech that we ensure all individuals are able to put forward their point of view about laws that they may be legitimately coerced into obeying. He is very concerned with the evangelist photographer's right to say that she does not want to take photographs at a same sex wedding. Yet he simultaneously treats differently the concerns of a same sex couple who, when faced with genuine hate speech, come to know through that hate speech that some people in society deem they are not worthy of equal treatment, and that their relationship is wrong, and who thereby become fearful of walking down the street holding hands, and fearful of violent attacks against them and their property on the basis of their sexuality. Could not such experiences, and the wider knowledge of those experiences among all same-sex attracted people, mean that the same sex couple could be silenced in much the same way Weinstein's evangelical photographer felt unable to express her views before a law on same sex marriage was passed? Don't communities who know their members regularly experience hate speech often feel unable to express their views on proposed upstream laws for fear they, too, will become targets of hate speech? I believe that they do, and have undertaken research showing that hate speech properly understood is one of the mechanisms by which this inability to express their views comes about.⁹

9. Katharine Gelber & Luke McNamara, *Evidencing the Harms of Hate Speech*, 22 *SOC. IDENTITIES* 324 (2016).

Both Weinstein and Alex Brown¹⁰ bemoan the paucity of evidence on this point, the silencing effect of hate speech (as opposed to Weinstein's concern with the silencing effect of hate speech laws). While there is not a great deal of such evidence, there are findings from psychology that show that individuals subjected to non-physical discrimination suffer significant harms to their physical and mental health.¹¹ These findings bolster and support the silencing claim, as do arguments in the literature about the indirect effects of hate speech, including the maintenance of power imbalances within social hierarchies of race.¹² My own, and my co-author's, empirical research interviewing members of communities targeted by hate speech shows that targets say they experience the harms of hate speech that are alleged in the literature.¹³ They testified to effects including that others were persuaded of negative stereotypes, a conditioning of the environment such that racism was normalized, subordination, silencing, fear, victimization, emotional symptoms, restrictions on freedom, lowering of self-esteem, maintenance of social imbalances of power, and undermining of their dignity. They testified that these harms were enduring. It is therefore not true that there is no evidence that silencing operates in the ways that defenders of hate speech laws allege.

10. Weinstein, *supra* note 1, at 575 (citing ALEXANDER BROWN, HATE SPEECH LAW: A PHILOSOPHICAL EXAMINATION (2015))

11. I focus here on the sources that pay particular attention to hate speech as a form of non-physical discrimination: Ilan H. Meyer, *Prejudice, Social Stress and Mental Health in Lesbian, Gay and Bisexual Populations: Conceptual Issues and Research Evidence*, 129 PSYCHOL. BULL., 674 (2003); Erik Vijlerveld et al., *The Cortisol Response to Anticipated Intergroup Interactions Predicts Self-Reported Prejudice*, 7 PLOS ONE, no. 3, Mar. 2012, at 1; Kathryn Freeman Anderson, *Diagnosing Discrimination: Stress from Perceived Racism and the Mental and Physical Health Effects*, 83 SOC. INQUIRY 55 (2013); Yin Paradies et al., *Racism as a Determinant of Health: A Protocol for Conducting a Systematic Review and Meta-Analysis*, 2 SYS. REVS. 85 (2013); Gilbert C. Gee, *A Multilevel Analysis of the Relationship Between Institutional and Individual Racial Discrimination and Health Status*, 92 AM. J. PUB. HEALTH 615 (2002); Ricci Harris et al., *Racism and Health: The Relationship Between Experience of Racial Discrimination and Health in New Zealand*, 63 SOC. SCI. & MED. 1428 (2006); VICTORIAN HEALTH PROMOTION FOUND., MENTAL HEALTH IMPACTS OF RACIAL DISCRIMINATION IN VICTORIAN ABORIGINAL COMMUNITIES (2012).

12. M. R. ALLBROOK, STRATEGIES FOR CHANGE: ANTI-RACISM INITIATIVES FROM AUSTRALIA, CANADA, THE UNITED STATES OF AMERICA, AND THE UNITED KINGDOM (2001); Barbara Bloch & Tanja Dreher, *Resentment and Reluctance: Working With Everyday Diversity and Everyday Racism in Southern Sydney*, 30 J. INTERCULTURAL STUD. 193 (2009); Kevin M. Dunn & Jacqueline Nelson, *Challenging the Public Denial of Racism for a Deeper Multiculturalism*, 32 J. INTERCULTURAL STUD. 587 (2011).

13. Gelber & McNamara, *supra* note 9.

THE NEED TO ENSURE MAXIMUM POLITICAL PARTICIPATION

This brings me to my second point, which is Weinstein's emphasis on political participation. He emphasises how important it is to ensure that opportunities exist "for each individual to participate as an equal in the public conversation about society's collective decisions."¹⁴ He reiterates that "equal opportunity to participate in the political process, including in the public discussion of collective decisions, is essential to political legitimacy,"¹⁵ and that "laws about which there can be reasonable disagreement are subject to being rendered illegitimate if people capable of self-government are denied the equal opportunity to participate in the process by which they are enacted."¹⁶ I could not agree more. The requirement of equal opportunity to participate in the decision making that affects one's life is central to democratic legitimacy.

The quarrel I have is with how Weinstein has applied this precept; specifically in his failure to acknowledge what would happen to his argument if he were to concede that hate speech itself is capable of undermining the equal opportunity in decisionmaking that he and I agree is fundamental to political legitimacy.

Where hate speech itself is capable of undermining those opportunities, and where it can be identified as doing so, it would be a blinkered view of democratic legitimacy that would consciously ignore this problem. I have argued this point previously,¹⁷ integrating a Nussbaumian capabilities-approach to support a strong protection for a broad range of speech with a Habermasian construction of public discourse, the discourse that Weinstein¹⁸ and Robert Post¹⁹ place at the centre of free speech protection. This produces a robust argument for the protection of speech on public discourse grounds, based on a recognition that

14. Weinstein, *supra* note 1, at 528.

15. *Id.* at 536.

16. *Id.* at 539.

17. Katharine Gelber, *Freedom of Political Speech, Hate Speech and the Argument from Democracy: the Transformative Contribution of Capabilities Theory*, 9 CONTEMP. POL. THEORY 304 (2010).

18. James Weinstein, *Participatory Democracy as the Central Value of American Free Speech Doctrine*, 97 VA. L. REV. 491 (2011).

19. Robert Post, *Participatory Democracy and Free Speech*, 97 VA. L. REV. 477, 483 (2011).

the individuals who take part in processes of legitimation become capable of doing so through their individual participation in speech-based activities. This view integrates an understanding of the role of speech in individual agency, with the role of speech in facilitating the social organisation that underpins democratic processes of legitimation.

At the same time, this view acknowledges that if the formation of individuals' capacities to become capable of constructing and implementing their own conception of the good life, and their ability to instantiate that by engaging in processes of democratic legitimation, were imperilled by some speech, there is an argument for the regulation (although certainly not an automatic presumption of the form that regulation should take)²⁰ of that harmful speech. This argument requires us to take seriously the risks to social justice of some speech. Importantly, it pertains only to that speech which is capable of imperilling democratic processes of legitimation in this way. Thus, an argument for the regulation of hate speech—defined carefully and in a confined way as speech capable of imperilling the very processes of democratic legitimation with which Weinstein, Post, I, and others are concerned—is rendered possible.

My argument rests on a conception of hate speech as speech that is directed at historically identifiable minorities;²¹ that targets them with speech that is harmful to their involvement in processes of democratic legitimation. This is because, in my view, the connection of hate speech with historically identifiable and systemic discrimination is key to its success in discursively enacting harm to a sufficient degree that it would imperil a target's ability to participate in the political decision making that affects them. This differentiates discursively discriminatory hate speech from speech that may offend someone, hurt their feelings, be vituperative, or contain epithets, some of which may be capable of the discriminatory harms of hate speech, but none of which is definitively capable of doing so unless other factors are in place.

20. Criminal prohibition is only one (and not my favoured) policy approach to the problem of hate speech. Other approaches include the civil regulation of hate speech as an act of discrimination, or providing mechanisms and resources to enable communities to challenge the messages of hate speech and facilitate their speech-based response, thereby overcoming the silencing effect.

21. The direction of such speech at historically identifiable minorities is key to its ability to harm, and not merely to offend, because such speech discursively enacts discrimination that is analogous to other forms of systemic discrimination.

This principle overcomes the problems of relying on vituperative speech, or on the use of epithets, as the defining feature of hate speech.

It also assists in understanding the problem Weinstein raises about having a law that prevents a racist landlord from calling would-be Pakistani tenants “cockroaches,” but does not prevent those would-be tenants from calling the (presumably Caucasian) landlord a “cockroach.”²² Weinstein’s steeping in the requirement under First Amendment jurisprudence to avoid viewpoint discrimination at all costs blinds him to the differential harms of these two events. Calling Pakistanis “cockroaches” is a racist term of abuse that likens a racial minority in a Western society to an animal that requires extermination. It therefore has a meaning and force that simply do not apply were the would-be tenants to call the landlord by the same epithet. Context, social power, and history matter in determining the harm that is occasioned in hate speech.

To be sure, the argument I make here introduces a different challenge in identifying the kinds of hate speech that are capable of harming in the ways that I have outlined. However, this is not in principle an insurmountable problem, as the type of hate speech able to be regulated would need to be sufficiently harmful to be capable of preventing its targets from participating in decisions about laws that affect them, to be targeted at historically identifiable minorities, and to occur within a social context within which systemic discrimination against that minority persists.²³

Even should Weinstein not agree with the argument I have just outlined, he may admit that it ought to be conceivable in principle that some hate speech ought to be considered capable of undermining individuals’ ability to express their own views in relation to upstream laws that the state will have the right to coerce them into obeying. If that is the case, then hate speech itself is capable of undermining the democratic legitimacy that Weinstein and I are both concerned with protecting.

22. Weinstein, *supra* note 1, at 542.

23. This is an issue on which Weinstein and I have conversed often, and it is an issue I am developing further in work-in-progress.

HOW HATE SPEECH LAWS WORK IN PRACTICE

The third point I wish to raise is Weinstein's use of anecdotal claims about the application of hate speech laws in countries around the world, and his claim that his evidence shows a "sufficiently large number of cases to show that most hate speech laws, whatever their intent,"²⁴ do not protect moderate forms of expressing one's opinion that are not designed to be captured by hate speech laws. My quarrel here is that Weinstein has not substantiated his case.

First, the total number of cases that Weinstein cites in the main text in which a claim of racial hatred (in some form) was substantiated is one in Holland, three in England, one in Scotland, two in Austria, one in Finland, one in France, one in Spain and one in Canada. This is a total of 11 cases, some of which by Weinstein's own admission did not involve "hate speech" legislation, but rather other public order types of provisions that were used to shut down what was perceived to be hate speech. He adds to his list four complaints in which a claim of racial hatred was not substantiated, including (in the main text),²⁵ two in France, one in Belgium and one in England. The total of cases to which he refers is therefore 15.²⁶ A "sufficiently large number of cases" to evidence widespread misapplication of hate speech laws this most certainly is not.

The first point is that, if Weinstein wishes to show the misapplication of hate speech laws, it would be helpful if he would limit his enquiry to hate speech laws, and not the misapplication of any other law. Jeremy Waldron also makes this point in his reply to Weinstein in this volume.

The second point to make here is that this total of 15 cases in the main text (30 with additional cases in footnotes) is a miniscule proportion of the entirety of hate speech complaints that are made across the globe every year. I do not believe anyone has calculated the global number of complaints, but just to give a small indication of how unrepresentative Weinstein's sample is, my and my co-author's recent study into the operation of hate

24. Weinstein, *supra* note 1, at 561.

25. Weinstein includes some additional cases in his lengthy footnotes. I have only counted here the ones that appear in the text of the article.

26. In the footnotes 124, 131, and 132, Weinstein also cites and discusses 15 further cases, making a total of 30 cases. See Weinstein, *supra* note 1, at 558 n. 124, 560-61 nn. 131-32.

speech laws in Australia over 25 years showed that the total number of complaints lodged in any given year in Australia was approximately 200-350.²⁷ Weinstein's examples (leaving aside the Dutch case which occurred in 1979), occurred between 2001 and 2015, with 10 of them occurring between 2001 and 2010. During the same period of 2001-2010 there were 2128 complaints lodged in Australia.

In 2010 alone, there were 3770 hate crimes reported in Germany, which included hate crimes, incitement to hatred, and propaganda offenses.²⁸ In 2014-2015, there were 30,991 racially motivated public order offences in England and Wales.²⁹

These figures render Weinstein's claim that a "sufficiently large" number of misapplications is occurring unsustainable in empirical terms. A far more in-depth and systemic study would be needed regarding the operation of hate speech laws in practice to sustain this point.

Weinstein does not only claim that the number of such misapplications is large enough to be worrying. He argues further that cases that are unsubstantiated are "undoubtedly" causing a chilling effect on free speech.³⁰ Again, there is insufficient evidence provided to make such a claim. The Australian study to which I have already referred has shown that, in spite of claims to the contrary, no evidence of a chilling effect was found in the context of 25 years of the operation of hate speech laws in Australia.³¹ Instead, Waldron's empirical intuition—that hate speech laws by and large operate to prevent harmful hate speech, but not to stop people debating in non-vilifying ways on matters of public policy—was strongly upheld by the evidence obtained in that study.³²

Any law will have some misapplications, this is a regular feature of a range of laws including defamation laws, and the law

27. Katharine Gelber & Luke McNamara, *The Effects of Civil Hate Speech Laws: Lessons from Australia*, 49 L. & SOC. REV. 631, 641-42 (2015).

28. See OSCE Office for Democratic Inst. and Human Rights, *Germany*, HATE CRIME REPORTING (Apr. 4, 2017, 10:36 AM), <http://hatecrime.osce.org/germany>.

29. HANNAH CORCORAN ET AL., HATE CRIME, ENGLAND AND WALES, 2014/15, at 1, 8 (2015).

30. Weinstein, *supra* note 1, at 559.

31. Gelber & McNamara, *supra* note 27, at 656-57.

32. *Id.*; Katharine Gelber & Luke McNamara, *Changes in the Expression of Prejudice in Public Discourse in Australia: Assessing the Impact of Hate Speech Laws on Letters to the Editor 1992-2010*, 20 AUSTL. J. HUM. RTS. 99 (2014).

2017]

DEFINITIONS & EVIDENCE

629

of torts. No-one suggests that the existence and occasional misapplication of defamation laws or of the law of torts means that the laws themselves are fundamentally flawed in the sense that they undermine the democratic legitimacy of other, especially anti-discrimination, laws.

In fact, I agree with Weinstein that some of the cases he mentions, in particular the cases in which Christians put forward their views about homosexuality, ought not to have been considered hate speech and ought not to have been prosecuted. However, I do not agree that this very small number of cases substantiates an argument that hate speech laws are misapplied with sufficient frequency to undermine Waldron's point that they are intended only to be applied to genuinely harmful hate speech.